

Electro-Optical System Analysis and Design

A Radiometry Perspective

Cornelius J. Willers

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Willers, Cornelius J.

Electro-optical system analysis and design: a radiometry perspective / Cornelius J Willers.

pages cm

Includes bibliographical references and index.

ISBN 978-0-8194-9569-3

1. Electrooptics. 2. Optical measurements. 3. Electrooptical devices--Design and construction. I. Title.

TA1750.W55 2013

621.381'045--dc23

2013002619

Published by

SPIE—The International Society for Optical Engineering

P.O. Box 10

Bellingham, Washington 98227-0010 USA

Phone: +1 360 676 3290

Fax: +1 360 647 1445

Email: spie@spie.org

Web: <http://spie.org>

Copyright © 2013 Society of Photo-Optical Instrumentation Engineers (SPIE)

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

The content of this book reflects the work and thought of the author(s).

Every effort has been made to publish reliable and accurate information herein, but the publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Cover image “Karoo Summer,” by Fiona Ewan Rowett (www.fionarowett.co.za), used with permission.

Printed in the United States of America.

First printing



Preface

*If you have an apple and I have an apple
and we exchange apples,
then you and I will still each have one apple.
But if you have an idea and I have an idea
and we exchange these ideas,
then each of us will have two ideas.*
George Bernard Shaw

On Sharing

Teachers cross our paths in life. Some teachers have names, others leave their marks anonymously. Among my teachers at the Optical Sciences Center at the University of Arizona were James Palmer, Eustace Dereniak, and Jack Gaskill. They freely shared their knowledge with their students. Some teachers teach through the pages of their books, and here I have to thank Bill Wolfe, George Zissis, and many more. Many years ago, R. Barry Johnson presented a short course which influenced my career most decisively.

The intent with this book is to now share some of my experience, accumulated through years of practical radiometry: design, measurements, modeling, and simulation of electro-optical systems. The material presented here builds upon the foundation laid at the Optical Sciences Center. I had the opportunity to share this material in an academic environment at graduate level in an engineering school, thereby clarifying key concepts. Beyond the mathematics and dry theory lies a rich world full of subtle insights, which I try to elucidate. May this book help you, the reader, grow in insight and share with others.

Reductionism, Synthesis, and Design

The reductionist approach holds the view that an arbitrarily complex system can be understood by reducing the system to many, smaller systems

that can be understood. This view is based on the premise that the complex system is considered to be the sum of its parts, and that by understanding the parts, the sum can be understood. While the reductionist approach certainly has weaknesses, this approach works well for the class of problems considered in this book. The methodology followed here is to develop the theory concisely for simple cases, developing a toolset and a clear understanding of the fundamentals.

The real world does not comprise loose parts and simple systems. Once the preliminaries are out the way, we proceed to consider more complex concepts such as sensors, signatures, and simple systems comprising sources, a medium, and a receiver. Using these concepts and the tools developed in this book, the reader should be able to design a system of any complexity. Two concurrent themes appear throughout the book: fragmenting a complex problem into simple building blocks, and synthesizing (designing) complex systems from smaller elements. In any design process, these two actions take place interactively, mutually supporting each other. In this whirlpool of analysis and synthesis, uncontrolled external factors (e.g., the atmosphere, noise) influence the final outcome. This is where the academic theory finds engineering application in the real world. This book aims to demonstrate how to proceed along this road.

Toward the end of the book, the focus shifts from a component-level view to an integrated-system view, where the 'system' comprises a (simple or composite) source, an intervening medium, and a sensor. Many real-world electro-optical applications require analysis and design at this integrated-system level. Analysis and design, as a creative synthesis of something new, cannot be easily taught other than by example. For this purpose several case studies are presented. The case studies are brief and only focus on single aspects of the various designs. Any real design process would require a much more detailed process, beyond the scope of this book.

General Comments

The purpose with this book is to enable the reader to find solutions to real-world problems. The focus is on the application of radiometry in various analysis and design scenarios. It is essential, however, to build on the foundation of solid theoretical understanding, and gain insight beyond graphs, tables and equations. Therefore, this book does not attempt to provide an extensive set of ready-to-use equations and data, but rather strives to provide insight into hidden subtleties in the field. The atmosphere provides opportunity for a particularly rich set of intriguing observations.

The strict dictionary definition of ‘radiometry’ is the measurement of optical flux. In this book, the term ‘radiometry’ is used in its wider context to specifically cover the calculation of flux as well. This wider definition is commonly used by practitioners in the field to cover all forms of manipulation, including creation, measurement, calculation, modeling, and simulation of optical flux. The focus of this book is not on radiometric measurement but on the analysis and modeling of measured data, and the design of electro-optical systems.

Antoine de Saint-Exupéry once wrote, “You know you’ve achieved perfection in design, not when you have nothing more to add, but when you have nothing more to take away.” The painful aspect of writing a book is to decide what *not* to include. This book could contain more content on radiometric measurement, emissivity measurement, properties of different types of infrared detectors, or reference information on optical material properties; however, these topics are already well covered by other excellent books, much better than can be achieved in the limited scope of this book.

The book provides a number of problems, some with worked solutions. The scope of problems in the early chapters tend to be smaller, whereas the problems in later chapters tend to be wider in scope. The more-advanced problems require numerical solutions. Although it is certainly possible to read the book without doing the advanced problems, the reader is urged to spend time mastering the skills to do these calculations. This investment will pay off handsomely in the future. Some of the problems require data not readily found in book format. The data packages are identified (e.g., DP01) and are obtainable from the pyradi website (see Section D.3.4).

To the uninitiated, the broader field of radiometry is dangerous territory, with high potential for errors and not-so-obvious pitfalls. Our work in the design labs, on field measurement trials, and in the academic environment led to the development of a set of best practices, called the ‘Golden Rules,’ which strives to minimize the risk error. Some of these principles come from James Palmer’s class, while most were stripes hard earned in battle. The readers are urged to study, use, and expand these best practices in their daily work. Any feedback, on the golden rules or any other aspect of the book, would be appreciated.

A book is seldom the work of one mind only; it is the result of a road traveled with companions. Along this road are many contributors, both direct and inadvertent. My sincere thanks to all who made their precious time and resources available in this endeavor. My sincere thanks goes to

Riana Willers for patience and support, as co-worker on our many projects — her light footprints fall densely on *every single page* in this book: advising, scrutinizing every detail, debating symbols and sentences, editing text and graphics, compiling the nomenclature and index, and finally, acting as chapter contributor. Riana is indeed the ghost writer of this book! Fiona Ewan Rowett for permission to use her exquisite “Karoo Summer” on the front cover. The painting beautifully expresses not only the hot, semi-arid Karoo plateau in South Africa, but also expresses radiated light and vibrant thermal energy, the subject of this book. My teachers at the Optical Sciences Center who laid the early foundation for this work. Ricardo Santos and Fábio Alves for contributing to the chapter on infrared detector theory and modeling. The pyradi team for contributing their time toward building a toolkit of immense value to readers of this book. Derek Griffith for the visual and near-infrared reflectance measurements. Hannes Calitz for the spectral measurements, and Azwitamisi Mudau for the imaging infrared measurements. Dr Munir Eldesouki from KACST for permission to use the Bunsen flame measured data in the book. The many colleagues, co-workers, and students at Kentron (now Denel Dynamics), the CSIR, KACST, and the University of Pretoria for influencing some aspect of the book. Scott McNeill and Tim Lamkins for patience and guiding me through the publication process. Scott’s untiring patience in detailed correction deserves special mention. Eustace Dereniak for encouraging me to submit the book for publication. Barbara Grant, Eustace Dereniak and an anonymous reviewer for greatly influencing the book in its final form. Finally, Dirk Bezuidenhout, and the CSIR for supporting the project so generously in the final crucial months before publication.

Mark Twain wrote that he did not allow his schooling to get in the way of his education. It is my wish that you, my esteemed reader, will delve beyond these written words into the deeper insights. Someone else said that the art of teaching is the art of assisting in discovery. May you discover many rich insights through these pages.

Nelis Willers
Hartenbos
March 2013

Contents

Nomenclature	xvii
Preface	xxiii
1 Electro-Optical System Design	1
1.1 Introduction	1
1.2 The Principles of Systems Design	2
1.2.1 Definitions	2
1.2.2 The design process	2
1.2.3 Prerequisites for design	3
1.2.4 Product development approaches	4
1.2.5 Lifecycle phases	4
1.2.6 Parallel activities during development	7
1.2.7 Specifications	8
1.2.8 Performance measures and figures of merit	10
1.2.9 Value systems and design choices	11
1.2.10 Assumptions during design	11
1.2.11 The design process revisited	12
1.3 Electro-Optical Systems and System Design	14
1.3.1 Definition of an electro-optical system	14
1.3.2 Designing at the electro-optical-system level	15
1.3.3 Electro-optical systems modeling and simulation	16
1.4 Conclusion	17
Bibliography	17
2 Introduction to Radiometry	19
2.1 Notation	19
2.2 Introduction	20
2.3 Radiometry Nomenclature	23
2.3.1 Definition of quantities	23
2.3.2 Nature of radiometric quantities	25
2.3.3 Spectral quantities	25
2.3.4 Material properties	27
2.4 Linear Angle	27

2.5	Solid Angle	28
2.5.1	Geometric and projected solid angle	28
2.5.2	Geometric solid angle of a cone	29
2.5.3	Projected solid angle of a cone	31
2.5.4	Geometric solid angle of a flat rectangular surface	32
2.5.5	Projected solid angle of a flat rectangular surface	32
2.5.6	Approximation of solid angle	33
2.5.7	Projected area of a sphere	34
2.5.8	Projected solid angle of a sphere	35
2.6	Radiance and Flux Transfer	35
2.6.1	Conservation of radiance	35
2.6.2	Flux transfer through a lossless medium	37
2.6.3	Flux transfer through a lossy medium	38
2.6.4	Sources and receivers of arbitrary shape	38
2.6.5	Multi-spectral flux transfer	39
2.7	Lambertian Radiators and the Projected Solid Angle	41
2.8	Spatial View Factor or Configuration Factor	43
2.9	Shape of the Radiator	44
2.9.1	A disk	44
2.9.2	A sphere	45
2.10	Photometry and Color	45
2.10.1	Photometry units	45
2.10.2	Eye spectral response	46
2.10.3	Conversion to photometric units	47
2.10.4	Brief introduction to color coordinates	48
2.10.5	Color-coordinate sensitivity to source spectrum	49
	Bibliography	51
	Problems	53
3	Sources	57
3.1	Planck Radiators	57
3.1.1	Planck's radiation law	60
3.1.2	Wien's displacement law	62
3.1.3	Stefan–Boltzmann law	63
3.1.4	Summation approximation of Planck's law	64
3.1.5	Summary of Planck's law	65
3.1.6	Thermal radiation from common objects	65
3.2	Emissivity	65
3.2.1	Kirchhoff's law	69
3.2.2	Flux transfer between a source and receiver	70
3.2.3	Grey bodies and selective radiators	71
3.2.4	Radiation from low-emissivity surfaces	73
3.2.5	Emissivity of cavities	74

3.3	Aperture Plate in front of a Blackbody	75
3.4	Directional Surface Reflectance	75
3.4.1	Roughness and scale	76
3.4.2	Reflection geometry	77
3.4.3	Reflection from optically smooth surfaces	77
3.4.4	Fresnel reflectance	78
3.4.5	Bidirectional reflection distribution function	80
3.5	Directional Emissivity	83
3.6	Directional Reflectance and Emissivity in Nature	85
3.7	The Sun	86
	Bibliography	87
	Problems	91
4	Optical Media	97
4.1	Overview	97
4.2	Optical Mediums	98
4.2.1	Lossy mediums	98
4.2.2	Path radiance	99
4.2.3	General law of contrast reduction	102
4.2.4	Optical thickness	103
4.2.5	Gas radiator sources	103
4.3	Inhomogeneous Media and Discrete Ordinates	104
4.4	Effective Transmittance	105
4.5	Transmittance as Function of Range	108
4.6	The Atmosphere as Medium	108
4.6.1	Atmospheric composition and attenuation	108
4.6.2	Atmospheric molecular absorption	111
4.6.3	Atmospheric aerosols and scattering	112
4.6.4	Atmospheric transmittance windows	116
4.6.5	Atmospheric path radiance	118
4.6.6	Practical consequences of path radiance	120
4.6.7	Looking up at and looking down on the earth	121
4.6.8	Atmospheric water-vapor content	121
4.6.9	Contrast transmittance in the atmosphere	124
4.6.10	Meteorological range and aerosol scattering	127
4.7	Atmospheric Radiative Transfer Codes	129
4.7.1	Overview	129
4.7.2	MODTRAN TM	129
	Bibliography	130
	Problems	133
5	Optical Detectors	135
5.1	Historical Overview	135

5.2	Overview of the Detection Process	136
5.2.1	Thermal detectors	136
5.2.2	Photon detectors	138
5.2.3	Normalizing responsivity	140
5.2.4	Detector configurations	140
5.3	Noise	140
5.3.1	Noise power spectral density	141
5.3.2	Johnson noise	142
5.3.3	Shot noise	143
5.3.4	Generation–recombination noise	144
5.3.5	$1/f$ noise	145
5.3.6	Temperature-fluctuation noise	145
5.3.7	Interface electronics noise	146
5.3.8	Noise considerations in imaging systems	146
5.3.9	Signal flux fluctuation noise	146
5.3.10	Background flux fluctuation noise	147
5.3.11	Detector noise equivalent power and detectivity	147
5.3.12	Combining power spectral densities	149
5.3.13	Noise equivalent bandwidth	149
5.3.14	Time-bandwidth product	150
5.4	Thermal Detectors	151
5.4.1	Principle of operation	151
5.4.2	Thermal detector responsivity	152
5.4.3	Resistive bolometer	155
5.4.4	Pyroelectric detector	157
5.4.5	Thermoelectric detector	159
5.4.6	Photon-noise-limited operation	161
5.4.7	Temperature-fluctuation-noise-limited operation	163
5.5	Properties of Crystalline Materials	163
5.5.1	Crystalline structure	164
5.5.2	Occupation of electrons in energy bands	165
5.5.3	Electron density in energy bands	166
5.5.4	Semiconductor band structure	169
5.5.5	Conductors, semiconductors, and insulators	170
5.5.6	Intrinsic and extrinsic semiconductor materials	171
5.5.7	Photon-electron interactions	174
5.5.8	Light absorption in semiconductors	176
5.5.9	Physical parameters for important semiconductors	179
5.6	Overview of the Photon Detection Process	179
5.6.1	Photon detector operation	179
5.6.2	Carriers and current flow in semiconductor material	179
5.6.3	Photon absorption and majority/minority carriers	180

5.6.4	Quantum efficiency	181
5.7	Detector Cooling	183
5.8	Photoconductive Detectors	187
5.8.1	Introduction	187
5.8.2	Photoconductive detector signal	187
5.8.3	Bias circuits for photoconductive detectors	189
5.8.4	Frequency response of photoconductive detectors	190
5.8.5	Noise in photoconductive detectors	191
5.9	Photovoltaic Detectors	193
5.9.1	Photovoltaic detector operation	193
5.9.2	Diode current–voltage relationship	196
5.9.3	Bias configurations for photovoltaic detectors	197
5.9.4	Frequency response of a photovoltaic detector	202
5.9.5	Noise in photovoltaic detectors	203
5.9.6	Detector performance modeling	207
5.10	Impact of Detector Technology on Infrared Systems	210
	Bibliography	212
	Problems	218
6	Sensors	221
6.1	Overview	221
6.2	Anatomy of a Sensor	221
6.3	Introduction to Optics	223
6.3.1	Optical elements	223
6.3.2	First-order ray tracing	225
6.3.3	Pupils, apertures, stops, and f -number	226
6.3.4	Optical sensor spatial angles	230
6.3.5	Extended and point target objects	232
6.3.6	Optical aberrations	232
6.3.7	Optical point spread function	235
6.3.8	Optical systems	236
6.3.9	Aspheric lenses	237
6.3.10	Radiometry of a collimator	238
6.4	Spectral Filters	240
6.5	A Simple Sensor Model	240
6.6	Sensor Signal Calculations	242
6.6.1	Detector signal	242
6.6.2	Source area variations	244
6.6.3	Complex sources	245
6.7	Signal Noise Reference Planes	245
6.8	Sensor Optical Throughput	248
	Bibliography	250
	Problems	250

7 Radiometry Techniques	255
7.1 Performance Measures	255
7.1.1 Role of performance measures	255
7.1.2 General definitions	256
7.1.3 Commonly used performance measures	257
7.2 Normalization	261
7.2.1 Solid angle spatial normalization	261
7.2.2 Effective value normalization	261
7.2.3 Peak normalization	262
7.2.4 Weighted mapping	263
7.3 Spectral Mismatch	264
7.4 Spectral Convolution	265
7.5 The Range Equation	267
7.6 Pixel Irradiance in an Image	268
7.7 Difference Contrast	271
7.8 Pulse Detection and False Alarm Rate	272
7.9 Validation Techniques	275
Bibliography	275
Problems	276
8 Optical Signatures	279
8.1 Model for Optical Signatures	279
8.2 General Notes on Signatures	283
8.3 Reflection Signatures	284
8.4 Modeling Thermal Radiators	285
8.4.1 Emissivity estimation	287
8.4.2 Area estimation	288
8.4.3 Temperature estimation	290
8.5 Measurement Data Analysis	292
8.6 Case Study: High-Temperature Flame Measurement	295
8.7 Case Study: Low-Emissivity Surface Measurement	295
8.8 Case Study: Cloud Modeling	297
8.8.1 Measurements	297
8.8.2 Model	298
8.8.3 Relative contributions to the cloud signature	300
8.9 Case Study: Contrast Inversion/Temperature Cross-Over	300
8.10 Case Study: Thermally Transparent Paints	301
8.11 Case Study: Sun-Glint	302
Bibliography	303
Problems	304
9 Electro-Optical System Analysis	309
9.1 Case Study: Flame Sensor	309

9.2	Case Study: Object Appearance in an Image	311
9.3	Case Study: Solar Cell Analysis	315
9.3.1	Observations	315
9.3.2	Analysis	316
9.4	Case Study: Laser Rangefinder Range Equation	321
9.4.1	Noise equivalent irradiance	321
9.4.2	Signal irradiance	322
9.4.3	Lambertian target reflectance	323
9.4.4	Lambertian targets against the sky	324
9.4.5	Lambertian targets against terrain	325
9.4.6	Detection range	326
9.4.7	Example calculation	326
9.4.8	Specular reflective surfaces	327
9.5	Case Study: Thermal Imaging Sensor Model	330
9.5.1	Electronic parameters	330
9.5.2	Noise expressed as D^*	331
9.5.3	Noise in the entrance aperture	331
9.5.4	Noise in the object plane	332
9.5.5	Example calculation	333
9.6	Case Study: Atmosphere and Thermal Camera Sensitivity	334
9.7	Case Study: Infrared Sensor Radiometry	337
9.7.1	Flux on the detector	337
9.7.2	Focused optics	339
9.7.3	Out-of-focus optics	342
9.8	Case Study: Bunsen Burner Flame Characterization	344
9.8.1	Data analysis workflow	345
9.8.2	Instrument calibration	346
9.8.3	Measurements	348
9.8.4	Imaging-camera radiance results	350
9.8.5	Imaging-camera flame-area results	352
9.8.6	Flame dynamics	353
9.8.7	Thermocouple flame temperature results	354
	Bibliography	355
	Problems	356
10	Golden Rules	365
10.1	Best Practices in Radiometric Calculation	365
10.2	Start from First Principles	365
10.3	Understand Radiance, Area, and Solid Angle	366
10.4	Build Mathematical Models	366
10.5	Work in Base SI Units	367
10.6	Perform Dimensional Analysis	367
10.7	Draw Pictures	368

10.8	Understand the Role of π	371
10.9	Simplify Spatial Integrals	371
10.10	Graphically Plot Intermediate Results	372
10.11	Follow Proper Coding Practices	372
10.12	Verify and Validate	372
10.13	Do It Right — the First Time!	373
	Bibliography	373
A	Reference Information	375
	Bibliography	383
B	Infrared Scene Simulation	385
B.1	Overview	385
B.2	Simulation as Knowledge-Management Tool	386
B.3	Simulation Validation Framework	386
B.4	Optical Signature Rendering	387
	B.4.1 Image rendering	391
	B.4.2 Rendering equation	393
B.5	The Effects of Super-Sampling and Aliasing	396
B.6	Solar Reflection, Sky Background, and Color Ratio	398
	Bibliography	401
C	Multidimensional Ray Tracing	403
D	Techniques for Numerical Solution	407
D.1	Introduction	407
D.2	The Requirement	407
D.3	Matlab [®] and Python [™] as Calculators	409
	D.3.1 Matlab [®]	410
	D.3.2 Numpy and Scipy	410
	D.3.3 Matlab [®] and Python [™] for radiometry calculations	410
	D.3.4 The pyradi toolkit	411
D.4	Helper Functions	411
	D.4.1 Planck exitance functions	412
	D.4.2 Spectral filter function	413
	D.4.3 Spectral detector function	415
D.5	Fully Worked Examples	417
	D.5.1 Flame sensor in Matlab [®]	417
	D.5.2 Flame detector in Python [™]	421
	D.5.3 Object appearance in an image in Python [™]	424
	D.5.4 Color-coordinate calculations in Python [™]	430
	D.5.5 Flame-area calculation in Matlab [®]	434
	D.5.6 The range equation solved in Python [™]	435

D.5.7	Pulse detection and false alarm rate calculation . . .	436
D.5.8	Spatial integral of a flat plate in Matlab®	437
	Bibliography	440
E	Solutions to Selected Problems	441
E.1	Solid Angle Definition	441
E.2	Solid Angle Approximation	441
E.3	Solid Angle Application (Problem 2.4)	448
E.4	Flux Transfer Application	448
E.5	Simple Detector System (Problem 6.2)	450
E.6	InSb Detector Observing a Cloud (Problem 8.2)	451
E.7	Sensor Optimization (Problem 9.1)	459
F	Additional Reading and Credits	471
F.1	Additional Reading	471
F.2	Credits	471
	Bibliography	472
	Index	477

Chapter 1

Electro-Optical System Design

There are two ways of constructing a design.

One way is to make it so simple that there are obviously no deficiencies.

Another way is to make it so complicated that there are no obvious deficiencies.

Sir Charles Antony Richard Hoare

1.1 Introduction

Optical flux has a source and, for the applications considered in this book, also a destination (sometimes called a receiver or absorber). Having a source and a destination, it must also have a channel, path, or medium. The approach in this book is to consider all three components interacting with the flux. The presence of more than one component implies that the flux can be seen to operate in a system context, with elements of the system including at least a source, a medium, and a receiver. Accepting the notion of an electro-optical *system*, the system can be subjected to actions such as analysis, design, and testing.

The fundamental approach taken in this book is that an electro-optical system should be considered as a system with cause-and-effect implications. Although the components in the system may not interact in a physical or causal manner, the performance of the system can be expressed as a set of relationships. In these relationships the system's performance leads to interdependencies between parameters of the various components. For example, the maximum range performance of a laser rangefinder depends on laser power, atmospheric transmittance, and sensor noise all of which require tradeoff analysis to optimize the system. Hence, notwithstanding the autonomy of each component, from a system perspective, the design process induces a synthetic parameter interdependence between the various components in the system.

The premise for this approach asserts that electro-optical systems can be optimized by trading off parameters between different components of

the system. Such a capability provides freedom and power to optimize the system by appropriate design choices. Returning to the laser rangefinder, the cost and complexity of laser power can be traded against the cost and complexity of noise in the receiver: the selection of the appropriate low-noise design may ease the burden of higher-power laser technology.

This chapter provides an overview of the generic design process, thereby providing a basis for electro-optical system design. The basic philosophy outlined here has application in practically all design situations. The ideas presented come from system engineering principles, practical experience, and common sense. This chapter structures the activities that most designers use intuitively. Although it is true that simple systems can be handled intuitively, the design of complex systems requires stricter discipline. The thoughtless execution of a procedure can be dangerous, but a well-structured approach provides a better basis for a sound design.

1.2 The Principles of Systems Design

1.2.1 Definitions

A *system* is a collection of items required to perform a task in order to achieve a goal. All lower-level items that may influence the goal, *however remote*, should be included as part of the system. Thus, when considering the system as a whole, all influences and effects should be accounted for, and the design can therefore be optimal within the constraints of that particular system.

System engineering is the entire management process required to bring about and maintain a complex system. In this chapter, particular emphasis is placed on specification practices and design analysis, as these lie at the very core of electro-optical system design.

A *subsystem* or *system segment* is a small system in itself, usually with a common goal, common technology, or cause-and-effect theme. There is a hierarchy of levels (such as system, subsystem, and subsystem) to signify levels of sub-ness. The word 'system' is used to indicate something consisting of a number of smaller items and therefore applies to all levels. Where necessary, the terms 'subsystem' or 'greater system' are used to denote level.

1.2.2 The design process

Solution design can be defined as the synthesis of smaller entities contributing to the whole of a solution. Design is an art, and a brief definition

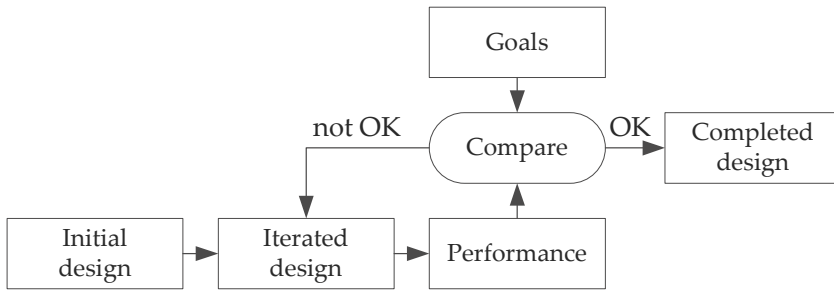


Figure 1.1 The basic design process.

cannot capture the subtleties inherent in the design process or the requirements for a good design. However, it may be enlightening to attempt a definition. The definition will be given in two phases: a short version in this section, and a more-complete version in Section 1.2.11.

Any design should start with a clear goal to be achieved and an initial design or design concept. The design goal can be translated into performance measures (see Section 1.2.8) by predicting performance in calculation, simulation, or by building the item. Once a system is built, the measured performance is compared to the design goals. If the measured performance compares acceptably with the goals, the design is completed. If the goals are not met by the design, the design details are iterated to find a better design, mostly by improving certain aspects of the current design. In certain cases the performance may not meet the system requirements but is acceptable to the customer, in which case the user requirement and system specifications are altered. This process is summarized in Figure 1.1.

1.2.3 Prerequisites for design

There are certain prerequisites that have to be satisfied before any design work can start. This may sound ludicrous, but these obvious prerequisites are often not met! The most important are:

1. There must be a concisely defined need. The developer must often help the client express the user-need in practical system requirements. Retrospectively, the client must also clearly understand the content of the developer's offer.
2. There must be adequate, approved resources in the form of labor, funding, schedule, and infrastructure.
3. There must be a knowledge base and a capability to design, develop,

test, and produce the system.

4. There must be a well-established technological base, including availability and capability, to support the development.
5. There must be a clear expression of mutual expectations between the client and developer. These expectations cover the deliverable product, the scope and nature of mutual support, the definition of contract completion, and delivery.

1.2.4 Product development approaches

Two different approaches are used in developing one-off versus mass-produced products. For mass-produced products, the primary goal during development is the compilation of *build and test documentation* suitable for use by the production line. The hardware products built during the development phases are mainly used to verify the design and documentation. An example of this situation is the volume production of products such as vehicles. For one-off product development, the requirement for documentation may be just as strict, but the product is proof of design and, at the same time, also the final deliverable product. An example of a one-off product is a large telescope, where each instance of the product is essentially unique.

1.2.5 Lifecycle phases

The lifecycle of a complex system can be divided into several distinct phases, each with a unique goal. It is difficult to isolate any one phase at any one instant because the phases flow into each other. If the formal system-engineering procedures are strictly applied, each phase culminates in a design review where the results are reviewed in terms of the goals for that phase. The decisions taken during the design review may lead to a change in strategy or system definition.

Not all of the phases shown here are always present. The need for a development phase depends on the complexity of the system and the maturity of similar systems. For example, a rework of an earlier design does not require many of the initial phases.

The various product lifecycle phases shown in Figure 1.2 can be summarized as follows:

Concept study: The concept study is a high-level study into different concepts that may be employed to achieve goals. Extensive use is made

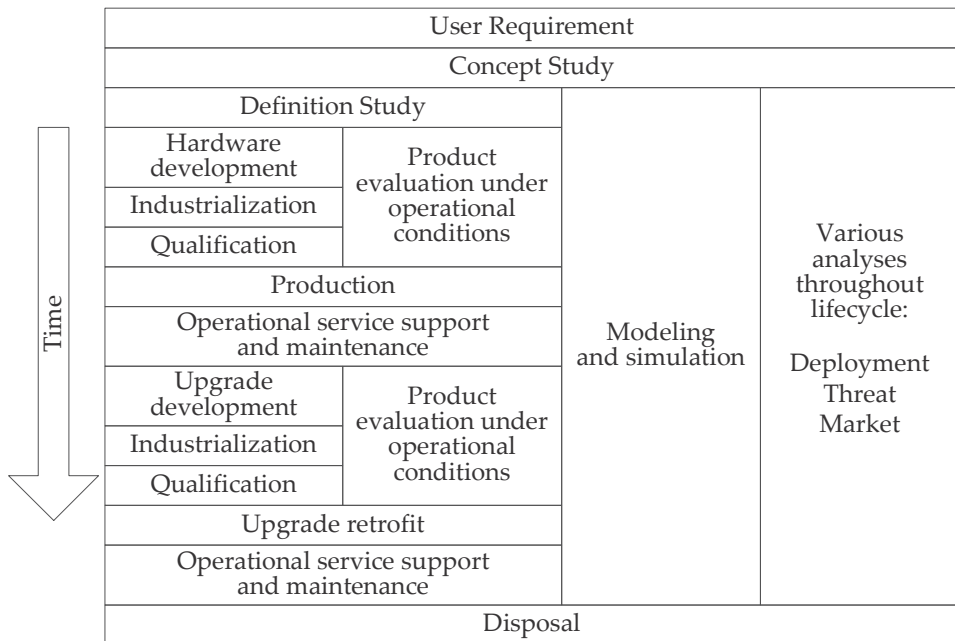


Figure 1.2 Development phases for complex systems.

of experienced, knowledgeable people to express opinions and do feasibility studies. These studies are concerned with functionality and not with implementation details. At the end of this phase the concepts, building blocks, and basic technologies required in the system are identified.

Definition study: During the definition study, the overall system is divided into subsystems or system segments. The subsystem functional requirements are then further allocated to hardware or software development requirements by trading off the required performance of the various items in the subsystem. In large projects, a formalized technical delegation structure must be developed. The structure depends on the work contents and available labor. The future development team will grow around this core of technical leaders. The primary aim is to develop design structure and quantify requirement specifications for the various subsystems and development items.

Hardware development, prototype: The aim of the prototype phase is to confirm that the hardware concept will indeed work. The hardware/software constructed at this time is not suitable for the final application, but it should perform all functions required (possibly at reduced performance). Considerable field tests are performed during

this phase. The work during this phase confirms the design approach at a conceptual level. Furthermore, valuable hardware experience is gained in the construction of the prototype.

Hardware development, experimental model: The hardware constructed in the experimental-model phase should be suitable for the final application, albeit with low confidence. All functions should be performed, and the size and packaging should be according to specification. Formal documentation-change control procedures are introduced to keep track of changes. The output from this phase comprises the design data pack, hardware, and software, designed and built to meet all specifications. In some cases the specifications are not yet met, and there may be some hardware and software functionality not yet finalized.

Hardware development, advanced model: Most of the design issues are cleared up toward the end of the advanced-model phase. The design documentation, test methods, and specifications are finalized. Unless required for specification conformity, the design will not change beyond this phase. During this phase the industrialization process is far advanced, but hardware might still be built in the development laboratories.

Industrialization: In the industrialization phase, the production processes are finalized by production specialists to ease future production. The design is only modified if necessary and then with approval from all concerned. Although this phase falls late in the development timeline, the industrialization personnel are involved in the design from the very beginning. In this phase, all hardware will be built on the production line.

System acceptance or qualification: System acceptance is the final evaluation testing where the system is evaluated against the requirement specification. After acceptance and approval by the user, full-scale production can begin.

Production, Operational Service Support, Maintenance, and Disposal:

These phases are part of the system lifecycle but are not covered under the present discussion.

Design reviews take place at predetermined times during the product lifecycle, most often at the end of each phase. The goals with these design reviews are to ascertain the design quality and applicability, assess performance risks, producibility, and several other aspects. There are formal design reviews where representatives from all affected interest groups

partake to assess the impact of the design on their own areas. Informal peer reviews are also very useful in that a quick answer can be obtained while working on a problem. Even though design reviews are sometimes regarded with apprehension and fear, they are very useful learning experiences for a well-prepared designer.

1.2.6 Parallel activities during development

Recall that Figure 1.2 shows four parallel activities during product development. The first activity, hardware development, was described in Section 1.2.5. This hardware development activity is supported by equally important but less-visible activities.

Product evaluation under operational conditions, e.g., during field trials, is an important reality check because it provides real-world feedback of the system's performance. Deficiencies and weaknesses can be identified, and limits of performance can be evaluated. Several operational-condition tests are typically executed during the development of a complex system. The nature of the operational tests shifts from initial exploration to evaluation later in the design phases, leading up to the qualification acceptance testing. Although laboratory tests are valuable, operational tests build confidence in the system.

Modeling and simulation provide several benefits; for example, it can provide a development environment for image processing algorithms, provide the capability to evaluate the system under conditions not possible in the real world, or evaluate different design options. It stands to reason that the models and simulation must be comprehensive and validated, relative to the task at hand. It is essential that the design in the simulation match the design in the hardware — there must be only one design, just different instances thereof.

Some system environments require ongoing analysis of elements in the external environment. One such example, the development of missile countermeasures, requires an ongoing analysis of the operation of missile threats. The outputs from these analysis tasks are used to influence some aspect of the design in order to best respond to the external environment.

To obtain maximum benefit from these four development activities, there should be a constant flow of information. Each of the activities should constantly re-evaluate its own position in the context of learning in the other activities. For example, the modeling and simulation activity should endeavor to most-accurately reflect the hardware status, and likewise, the hardware activity can learn from simulation experiments.

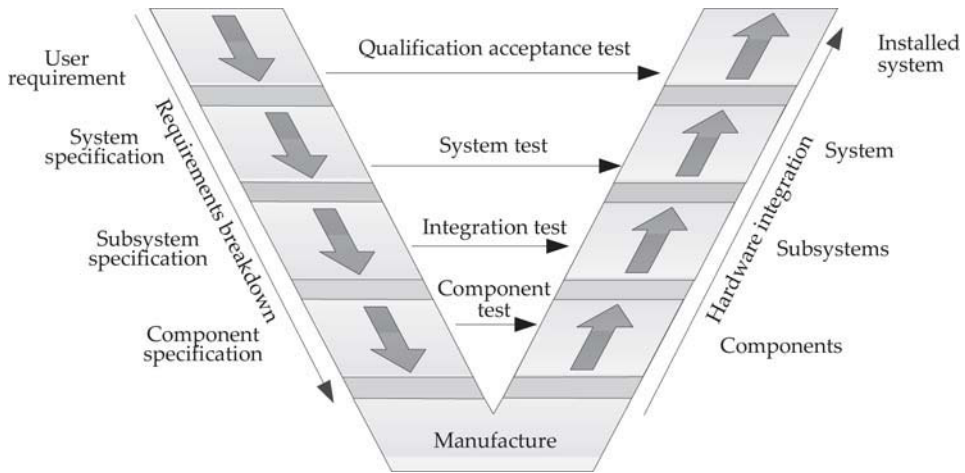


Figure 1.3 The systems engineering V-chart for system decomposition and integration.

1.2.7 Specifications

1.2.7.1 Requirement allocation and integration

One of the central themes in system engineering is the concept of a specification hierarchy. High-level requirements are interpreted, and requirements are allocated to lower-level specifications. The divide-and-conquer theory states that if the subsystems are designed to meet the lower-level specifications, the integrated hardware should be able to meet the higher-level specifications. Integration is the process of putting all items together to form a system at a higher level. This concept is demonstrated in the systems engineering V-chart,¹ shown here in Figure 1.3. Appropriate testing is done at each level, in accordance with the requirement at that level.

At any level there are *requirements* (or needs) and real or perceived *restrictions* that must be reconciled in the next-lower-level specifications. Requirements include aspects such as application environment (temperature, pressure, vibration, shock, and humidity), required functions, required performance, electrical and mechanical interfaces, and so forth. Restrictions include aspects such as system restrictions (size, weight, power, shape), technological limitations, and resource limitations (funding and labor). Restrictions also include requirements imposed by other subsystems, acting as restrictions on this subsystem. The main objective of specification allocation is to break down the large system specification into smaller, independently definable but interrelated specifications expressible in hardware items that can be laboratory tested and validated.

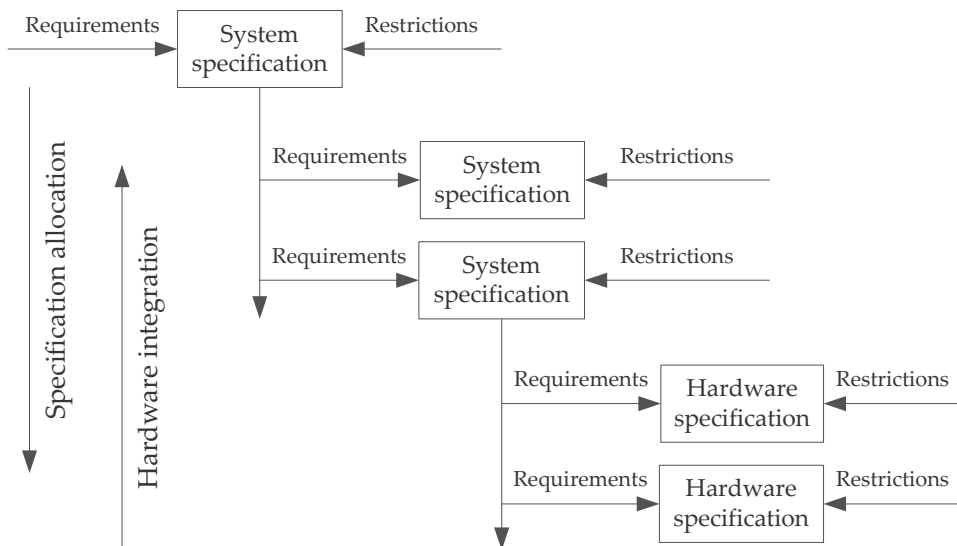


Figure 1.4 A schematic representation of the requirements allocation process.

The hierarchical allocation from higher to lower levels stops at any level where the specified item is a well-defined, self-contained item that can be conveniently managed and tested. The allocated specification should only contain critical requirements, allowing the designer freedom to explore alternative options during design. The hardware or software is then developed to meet the specification. In order to evaluate the performance of the total system, the individual hardware items are then integrated into larger systems, upward through the hierarchical specification structure. The hierarchical approach therefore allows the separation of needs and requirements during the initial phases, and also the integration of hardware in the advanced development phases.

The process of specification allocation and integration is never terminated. Throughout all development phases the process continues, albeit with different intensities. This continual process is due to the fact that new knowledge of the system is accumulated, leading to changes in definition and improved specifications. Sometimes new information reveals severe restrictions in the system as defined at that time. Sometimes it may be possible to change the system definition, but if the design is in an advanced stage, the changes may not be implemented.

The hierarchical allocation process is shown schematically in Figure 1.4. Note how the specification allocation to lower levels takes place in the downward direction, whereas hardware integration takes place from the bottom upward. It often happens that items are operational at a lower

level, but during the integration process, the higher-level system does not perform as expected. From the foregoing it is clear that the hierarchical breakdown provides an ordered approach for the design and integration of complex systems. However, the success of this approach depends totally on the discipline and diligence of the people employing these methods.

1.2.7.2 Specification practices

A specification for a subsystem or hardware item should be a complete, clear, and concise statement of the required functionality, minimum acceptable performance, and restrictions. A requirement that states that a system must operate under 'typical continental atmospheric conditions' is not a specification but a user need. The specification provides performance measures that can be tested against. All requirements should be given in measurable, laboratory-testable quantities. The specification should exist in version-controlled documents for purposes of future reference and communication. The specification should have sufficient information but not too much detail or over-specification (unnecessary or overly tight requirements). Finally, the specification should be kept current with the project development status. The practice of specification writing is well documented in system-engineering literature.¹

1.2.8 Performance measures and figures of merit

A figure of merit (FOM) or technical performance measure (TPM) is a means to express some aspect of the system's performance. Whereas a specification is a fixed contractual objective, the FOM/TPM provides measured or predicted information about the system. FOM/TPMs can be used at all levels of design to optimize the design at that particular level. Typical examples include signal-to-noise ratio (SNR) as a function of bandwidth, detection range under different conditions, or modulation transfer function (MTF) as a function of aperture diameter.

During the initial stages of design, a number of FOM/TPM measures are selected to act as indicators of system performance. The first build of a complex system does not always perform to the final specification. In such cases the development plan makes provision for poorer performance in initial stages but reaches the full performance at the end of development. The planned relative growth in performance is represented by allocating appropriate TPM values at different stages of development. During the various development stages, the designer measures the system's FOM/TPM measures and compares it against the planned values. If the system performs below the planned level, action plans are then initiated

to rectify the situation.

A FOM is typically influenced by a number of different system parameters. As a design tool, the sensitivity of the FOM to system parameters can be studied, and design trade-offs made.

1.2.9 Value systems and design choices

The designer is frequently faced with choices, and the best selection may not be clear. In this regard a *value system* is of great help. A value system is a hierarchical set of priorities and goals that helps the designer to make difficult choices, by evaluating design options in terms of predetermined priorities.

To compile the value system, list all of the important issues concerning the user need, the hardware to be designed, and the design environment. The aspects to be addressed should include the importance of development time scale, product cost, performance, local content, reliability, maintainability, upgradeability, etc. Any real value system is further complicated by peripheral but important issues such as company policy and mission, labor situation, profit motives, etc. The project value system should not be contaminated by ulterior motives such as personal gain, particular interests or disinterests, and 'old-boys' understandings.

When faced with design choices, the designer can allocate points for each entry in the value system, to each option. The option that best satisfies the high-priority values is then the desired option.

1.2.10 Assumptions during design

When faced with problems in an examination, students are often exhorted to make assumptions and continue to find a solution. Assumptions are part of our everyday life, and people who cannot make or accept assumptions are under severe stress. Our readiness to make assumptions numbs our consideration about the potential consequences of these assumptions. This is particularly true about technical assumptions because very few designers check all of their assumptions. The problem with assumptions lies in the fact that erroneous assumptions are only realized when it is too late, and the cost implications are potentially high. Unchecked assumptions can cause inadequate performance, catastrophic failures, and high redesign costs. Clearly, something must be done to prevent these situations.

The assumption confirmation procedure proposed here is very effec-

tive if applied with diligence and careful thought. The procedure is simple:

1. Write down *all* assumptions immediately after making them. Write down as many as you can think of; some may seem trivial now but may be important later. Collect all of the assumptions in a central database or document.
2. Prioritize the assumptions by three independent ratings (each on a three-point scale): severity or criticality, urgency, and risk. Add them up to derive the overall priority for that assumption.
3. Review and confirm the top-priority assumptions, and ignore the low-priority assumptions, as there is not enough time to do them all.
4. Take corrective actions if assumptions were shown to be incorrect.
5. Keep the complete list — never discard any assumptions, however trivial.
6. Ask a colleague or a design-review panel to go over the complete list of the assumptions at regular intervals to confirm your judgment.

1.2.11 The design process revisited

The previous sections describe some activities of the technological design process; integration of the various activities is depicted in Figure 1.5. On first appearance the figure seems unwieldy, but a little study reveals that it describes the intuitive process that takes place during design. Depending on the complexity of the design, certain functions may require a formal action, whereas others may take place in the mind of the designer without any formal manifestation.

As an example, consider the compilation of a model. The designer has a certain expectation of the required item performance, and changes the design until the expectations are fulfilled. If the item is small enough for the designer to maintain a mental model, there is no need to have the model in formal mathematical form. In more-complex systems, and in particular in electro-optical systems, the model can be quite complex, and a mathematical or numerical implementation is required.

Note the symmetry between modeling activities in the upper half of the diagram and hardware activities in the lower half of the diagram. There is a duality between the modeling and hardware activities, where modeling is used to derive a theoretical solution through analysis while hardware construction finds a physical solution. The modeling activities

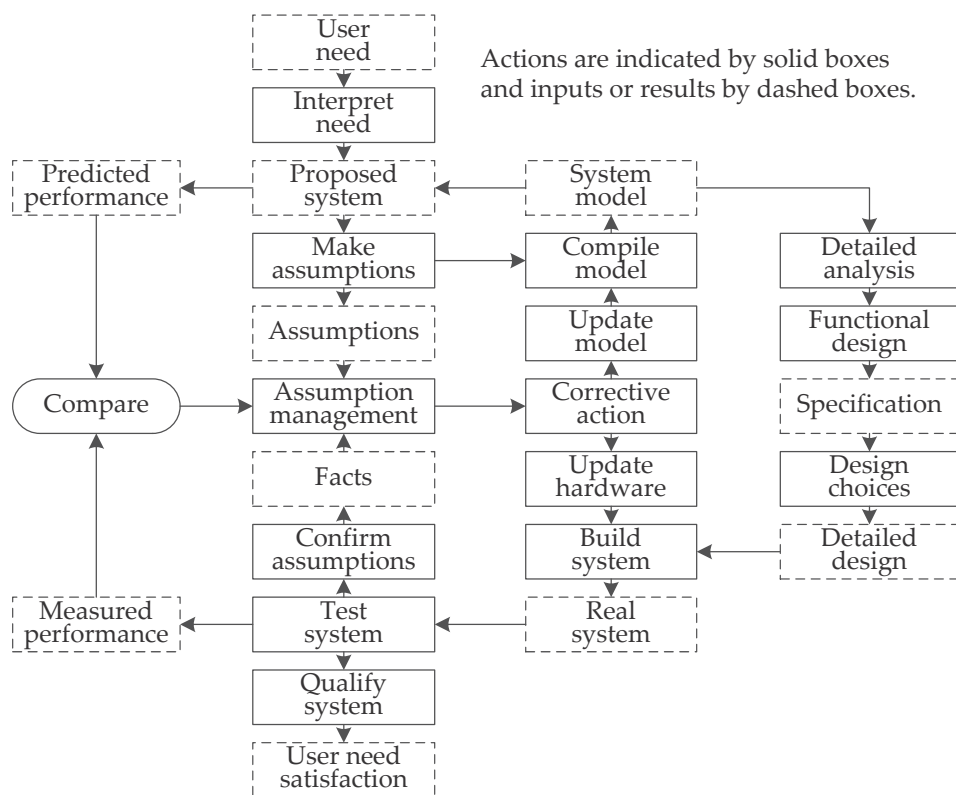


Figure 1.5 Actions and results during the design process.

normally start before the hardware design activities, but they continue while the design process takes place, as shown in Figure 1.2. A model is used to derive the hardware design details, and hardware evaluation tests are used to confirm and improve the model, in a never-ending cycle. The parallel development of models and hardware therefore greatly increases confidence in the system. This is especially true for systems where human operators or free atmospheric effects are part of the system.

It should be clear that system design is not a simple one-dimensional problem. There are many complex issues involved in an iterated design approach, frequently requiring several experimental models before the final design is approved. The design process normally stops when the design is certifiably demonstrated to meet the requirement. The designer is normally not allowed to have the final say in the acceptance of the design, but a review panel or advisors grants the final approval. It often happens that the design process continues into production if unexpected problems arises.

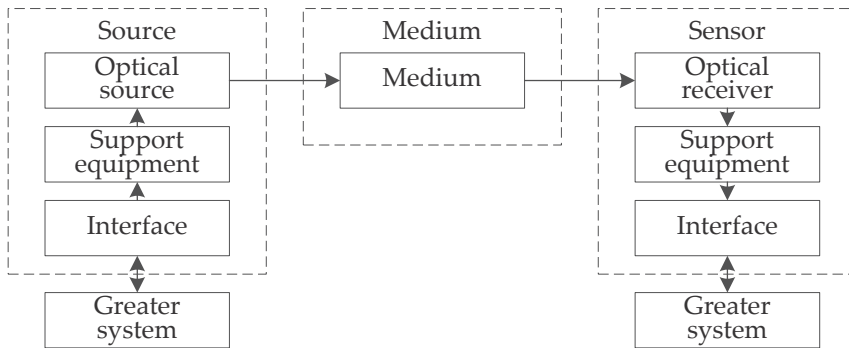


Figure 1.6 A generic electro-optical system.

1.3 Electro-Optical Systems and System Design

1.3.1 Definition of an electro-optical system

An *electro-optical system* is defined as a collection of items utilizing optical, physical, and electronic techniques to achieve some goal. Any physical object or medium that affects the flow of optical energy, or the conversion between optical and electronic energy, should be considered as part of the electro-optical system. A generic model of an electro-optical system is shown in Figure 1.6. The need for the separate identification of an electro-optical system is driven by the fact that the true effect of an item at the lowest level is often not fully appreciated. For instance, atmospheric turbulence may be a major factor when determining electronic time constants in a complex signal processor functionally far removed from the optical units. It is only when the integrated electro-optical system is analyzed that some subtle interdependencies can be handled in the design. Note that the human user is often a critical part of the system, usually as part of the receiver. This book enables the reader to perform electro-optical systems design at this higher, integrated level.

Three major components are present:

The Source, comprising the interface with the wider-scope system-support equipment, optical energy source, optics, and packaging. It also includes interference and spurious background sources.

The Medium, comprising all effects that in any way influence the optical signal. This may include optical fibers, atmospheric attenuation, dust, smoke, turbulence effects, etc.

The Sensor, comprising the packaging, optics, optical energy detector, sup-

Table 1.1 Examples of typical electro-optical systems cast into the generic format.

System	Heat-seeking missile	Fiber optic link	Free-space laser link
Source	The infrared flux from an aircraft is influenced by engine temperature, altitude, and flying conditions. The energy cannot be directed or contained in a beam, i.e., uncooperative. Hot objects such as flares & the sun act as false targets.	A laser or LED source is used. The coupling of the source energy into the optical fiber is very important, otherwise, the source presents no major problems to the designer.	A laser provides high-power pulses into a well-defined optical beam. The beam divergence is dictated by external considerations, such as pointing accuracy and atmospheric turbulence effects.
Medium	Atmosphere. Losses by scattering & absorption. Decreases with altitude. Strong plume signature attenuated by carbon dioxide. Clouds may cause obscuration of the target aircraft.	The fiber contains the energy. Impurities in the fiber absorb, and the molecular structure of the fiber scatters energy. The dispersion in the fiber causes multi-path transmission and mode dispersion.	Atmosphere. Losses by scattering & absorption. The laser beam divergence contains the energy within limits, but atmospheric turbulence affects beam quality by divergence, scintillation, and beam wander.
Sensor	Narrow field of view on a steerable, stabilized platform. Discriminates the aircraft from other hot objects in the scene. Interference and detector noise in the seeker limits sensor sensitivity. The severe size restriction limits design options.	Accepts optical signals from the fiber with little external interference or background flux. Electronic bandwidth is very high, leading to higher noise power.	Receiver beam width is determined by pointing accuracy and the background flux causing noise in the detector. Atmospheric turbulence necessitates the use of large receiving apertures to reduce scintillation modulation depth.

port equipment, signal conditioning, functional interface with the rest of the system, and any part of the greater system that may be affected by items in this system. The human may be tasked to detect, recognize, or identify objects in a scene.

Table 1.1 lists typical examples of electro-optical systems. Note how all of the systems fit into the basic source–medium–sensor model.

1.3.2 Designing at the electro-optical-system level

The typical activities of a non-trivial, electro-optical system design include work at the level of the individual components (source, medium, and sensor) but also at the integrated system level.

The source must be characterized either by measurement or design

choice. The signature of sources emitting optical flux must be measured, modeled, and understood. Active sources such as lasers must be designed to meet specific requirements. The requirement is that the properties of the source, whatever its nature, must be well understood and quantified.

The medium must be designed to requirement or characterized if it is not a design item (i.e., the atmosphere). It is important to quantify all key properties of the medium that may affect the system performance.

The sensor design is normally more directly under the designer's control than are the source and medium. The designer must select and optimize the optics, spectral filters, detectors, electronics, and packaging to meet the requirements. In some instances the sensor design must integrate tightly with other components, such as stabilized gimbal platforms, which places severe restrictions on the optical design.

At the level of an integrated system, most design and analysis focuses on system performance. The interdependency between a component's properties and its effect on performance are studied and optimized. The effect of the medium on system performance must be offset by appropriate design choices in the source or sensor. These effects are only quantifiable at the system level.

The advantages of designing at the integrated electro-optical system level include the following: (1) User requirements can be satisfied by trading off the various items present in the system. For example, the source output power can be reduced if the sensor can be made more sensitive. (2) The effective performance of the electro-optical system can only be determined in the context of the system because each item independently cannot perform the system function. (3) Certain key decisions can only be made at a system level because they influence all lower-level items. For example, the pulse width of a laser pulse effects the laser design as well as the sensor detector and electronics design.

1.3.3 Electro-optical systems modeling and simulation

The multi-disciplinary nature of electro-optical systems makes it very difficult to build a simple analytical model of the whole system. Comprehensive modeling and large-scale simulation^{2,3} are often done to create environments where system-level investigations and design can be executed (see Appendix B). Maximum benefit can be obtained from such simulation activities if they are tightly integrated with the hardware design process. System-level modeling and simulation aids the designer as follows:

1. Optimal specification allocation can be achieved by analyzing the in-

terdependencies between the various components in the system (see Section 1.2.7.1).

2. System performance can be predicted before hardware is built and for scenarios that may be not be easily tested in real life.
3. Operational evaluation can be simulated prior to the real-world event, investigating and confirming the test design prior to deployment.
4. In a sufficiently accurate simulation, the simulation provides a development environment, e.g., for image-processing algorithm or mode-control development.
5. Parallel development of the simulation and hardware by separate teams provides mutual insight and peer review to both teams' work.
6. The simulation provides an easy way to perform system-level tolerance analysis and error-budget analysis on allocated performance specifications, saving iterative hardware trials.

In large and complex systems, the system modeling and simulation effort can be as important to the end goals as the hardware design effort, and substantial time should be invested in this area.

1.4 Conclusion

The principles described in this chapter can significantly ease the design of complex systems with increased confidence in the system performance. In practice it is very seldom possible to apply all of the principles described above due to labor or schedule constraints. It is still useful to consider the complete process and apply certain aspects in a particular situation. An analysis of this nature is never complete, but the basic approach outlined here should be applicable to most electro-optical design situations.

Bibliography

- [1] IncoSE, "Systems Engineering Handbook," <http://www.incoSE.org/ProductsPubs/products/sehandbook.aspx>.
- [2] DIRSIG Team, "Digital Imaging and Remote Sensing Image Generation (DIRSIG)," <http://www.dirsig.org/>.
- [3] Willers, M. S. and Willers, C. J., "Key considerations in infrared simulations of the missile-aircraft engagement," *Proc. SPIE* **8543**, 85430N (2012) [doi: 10.1117/12.974801].

Chapter 2

Introduction to Radiometry

*He who loves practice without theory
is like the sailor who boards ship without a rudder and compass
and never knows where he may cast.*

Leonardo da Vinci

2.1 Notation

In this book the difference operator ‘ d ’ is used to denote ‘a small quantity of ...’. This ‘small quantity’ of one variable is almost always related to a ‘small quantity’ of another variable in some physical dependency. For example, irradiance is defined as $E = d\Phi/dA$, which means that a small amount of flux $d\Phi$ impinges on a small area dA , resulting in an irradiance of E . ‘Small’ is defined as the extent or domain over which the quantity, or any of its dependent quantities, does not vary significantly. Because any finite-sized quantity varies over a finite-sized domain, the d operation is only valid over an infinitely small domain $dA = \lim_{\Delta A \rightarrow 0} \Delta A$.

The difference operator, written in the form of a differential such as $E = d\Phi/dA$, is not primarily meant to mean differentiation in the mathematical sense. Rather, it is used to indicate something that can be integrated (or summed). In this book, differentiation is seldom done, whereas integration is done on almost every page.

In practice, it is impossible to consider infinitely many, infinitely small domains. Following the reductionist approach, any real system can, however, be assembled as the sum of a set of these small domains, by integration over the physical domain as in $A = \int dA$. Hence, the ‘small-quantity’ approach proves very useful to describe and understand the problem, whereas the real-world solution can be obtained as the sum of a set of such small quantities. In almost all of the cases in this book, it is implied that such ‘small-quantity’ domains will be integrated (or summed) over the (larger) problem domain.

Photon rates are measured in quanta per second. The ‘second’ is an SI unit, whereas quanta is a unitless count: the number of photons. Photon rate therefore has units of $[1/\text{s}]$ or $[\text{s}^{-1}]$. This form tends to lose track of the fact that the number of quanta per second is described. The book may occasionally contain units of the form $[\text{q}/\text{s}]$ to emphasize the photon count. In this case, the ‘q’ is not a formal unit, it is merely a reminder of ‘counts.’ In dimensional analysis (Section 10.6) the ‘q’ is handled the same as any other unit.

Unless otherwise stated, temperatures are in kelvin $[\text{K}]$ (note that the use of kelvin as temperature is without the word ‘degrees’).

2.2 Introduction

Electromagnetic radiation can be modeled as a number of different phenomena: rays, electromagnetic waves, wavefronts, or particles. All of these models are mathematically related. The appropriate model to use depends on the task at hand. Either the electromagnetic wave model (developed by Maxwell) or the particle model (developed by Einstein) are used when most appropriate. The electromagnetic spectrum is shown in Figure 2.1.

The electromagnetic wave energy is propagated as alternating, perpendicular, sinusoidal electric and magnetic fields [Figure 2.2(a)]. The electric and magnetic fields are related by Maxwell’s equations¹ such that a spatially varying electric field causes the magnetic field to change over time (and vice versa). The interaction between the two fields results in the propagation of the wave through space, away from the source. In the near field, close to the source, the wave diverges spatially, and there is a 90-deg phase shift between the electric and magnetic fields. In the far field, the wave has negligible divergence, and the magnetic and electric fields are in phase.

In terms of the wave model [Figure 2.2(c) and (d)], light can be considered a transverse wave propagating through a medium. The wave has a wavelength equal to the distance over which the wave’s shape repeats. The distance along a wave is also measured in phase, where the signal repeats once with every 2π rad phase increase (phase angle is derived from the rotating phasor model of a sinusoidal generator). The wave’s wavelength λ in $[\text{m}]$ is related to the frequency ν in $[\text{Hz}]$, by its propagation speed c in $[\text{m}/\text{s}]$ and the medium’s index of refraction n by $c = n\lambda\nu$. For light in vacuum, the speed c is a universal physical constant.

The *wavefront* is the wave’s spatial distribution at a given wave phase angle. An optical ray can be modeled as a vector such that it is perpen-

Electromagnetic Radiation

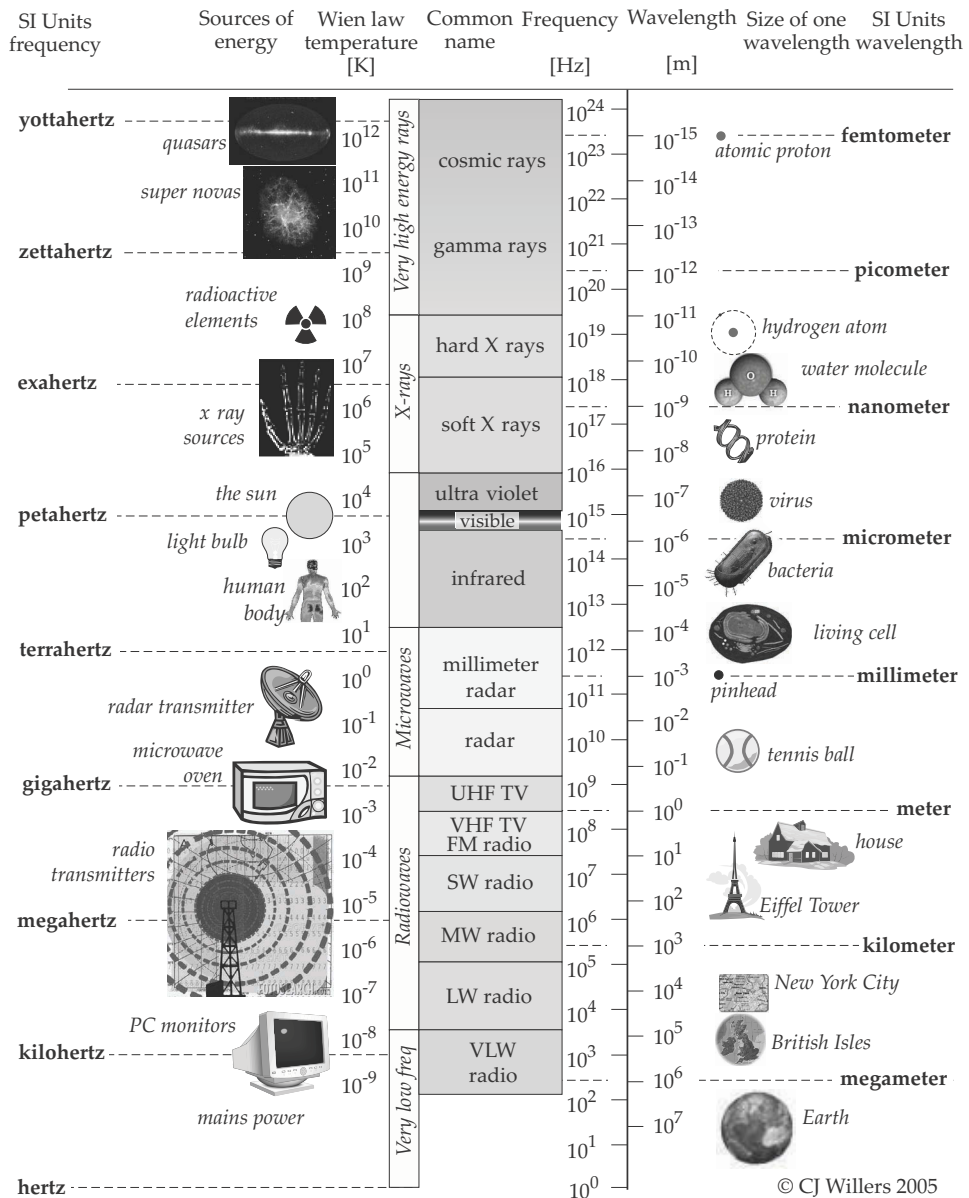


Figure 2.1 The electromagnetic spectrum.

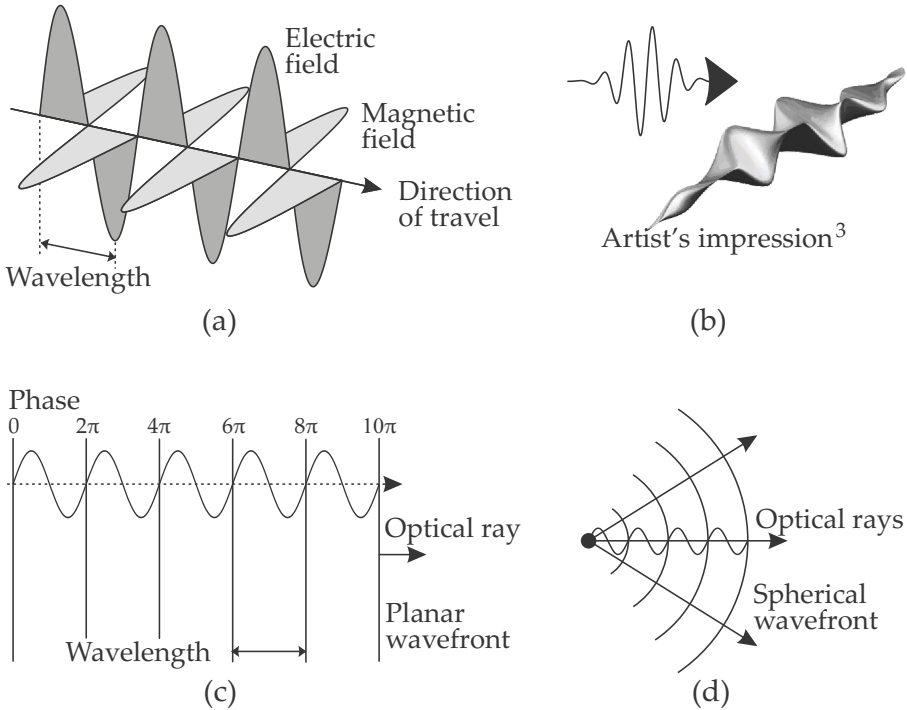


Figure 2.2 Simple light models: (a) electromagnetic wave, (b) photon wave packet, (c) far-field, plane wave, and (d) near-field spherical wave.

pendicular to the wavefront. The wavefront and ray models are mostly used in aberration calculations during optical design and when laying out an optical system.

The *photon*² is a massless elementary particle and acts as the force carrier for the electromagnetic wave. Photon particles have discrete energy quanta proportional to the frequency of the electromagnetic energy, $Q = h\nu = hc/\lambda$, where h is Planck's constant. Attempting to fuse the particle model and the wave model, consider the photon as a spatially limited electromagnetic disturbance (a wave packet) propagating through space. In this book the photon is sometimes graphically illustrated with the symbol shown in Figure 2.2(b). The figure also shows an artist's impression of the photon as a wave packet.

Radiometry is the measurement and calculation of electromagnetic flux transfer⁴⁻¹⁰ for systems operating in the spectral region ranging from ultraviolet to microwaves. Indeed, these principles can be applied to electromagnetic radiation of any wavelength. This book only considers ray-based⁹ radiometry for incoherent radiation fields.

2.3 Radiometry Nomenclature

2.3.1 Definition of quantities

The nomenclature is defined on the ISO/CIE and ANSI lighting standards^{11–14} and applied in various texts.^{4,7,9,10} Radiometric quantities are denoted by single-character symbols (Roman or Greek letters), such as Q , Φ , w , I , M , E , and L . These quantities are defined in Table A.1 and shown in Figure 2.3 (expanded from Pinson⁵) and Table 2.1. Radiometry nomenclature was revised during the 1960s and early 1970s with the result that some of the older textbooks¹⁵ are left using the older notation (P , W , N , H).

Radiometric quantities can be defined in terms of three different but related units: radiant power (watts), photon rates (quanta per second), or photometric luminosity (lumen). Photometry is radiometry applied to human visual perception.^{5,14,16–18} The conversion from radiometric to photometric quantities is covered in more detail in Section 2.10. It is important to realize that *the underlying concepts are the same, irrespective of the nature of the quantity*. All of the derivations and examples presented in this book are equally valid for radiant, photon, or photometric quantities.

Flux is the amount of optical power, a photon rate, or photometric luminous flux, flowing between two surfaces. There is always a source area and a receiving area, with the flux flowing between them. All quantities of flux are denoted by the symbol Φ . The units are [W], [q/s], or [lm], depending on the nature of the quantity.

Irradiance (areance) is the areal density of flux on the receiving surface area, as shown in Figure 2.3(a). The flux flows inward onto the surface with no regard to incoming angular density. All quantities of irradiance are denoted by the symbol E . The units are [W/m²], [q/(s·m²)], or [lm/m²], depending on the nature of the quantity.

Exitance (areance) is the areal density of flux on the source surface area, as shown in Figure 2.3(b). The flux flows outward from the surface with no regard to angular density. The exitance leaving a surface can be due to reflected light, transmitted light, emitted light, or any combination thereof. All quantities of exitance are denoted by the symbol M . The units are [W/m²], [q/(s·m²)], or [lm/m²], depending on the nature of the quantity.

Intensity (pointance) is the density of flux over solid angle, as shown in Figure 2.3(c). The flux flows outward from the source with no regard for surface area. Intensity is denoted by the symbol I . The human perception of a point source (e.g., a star at long range) ‘brightness’ is an intensity

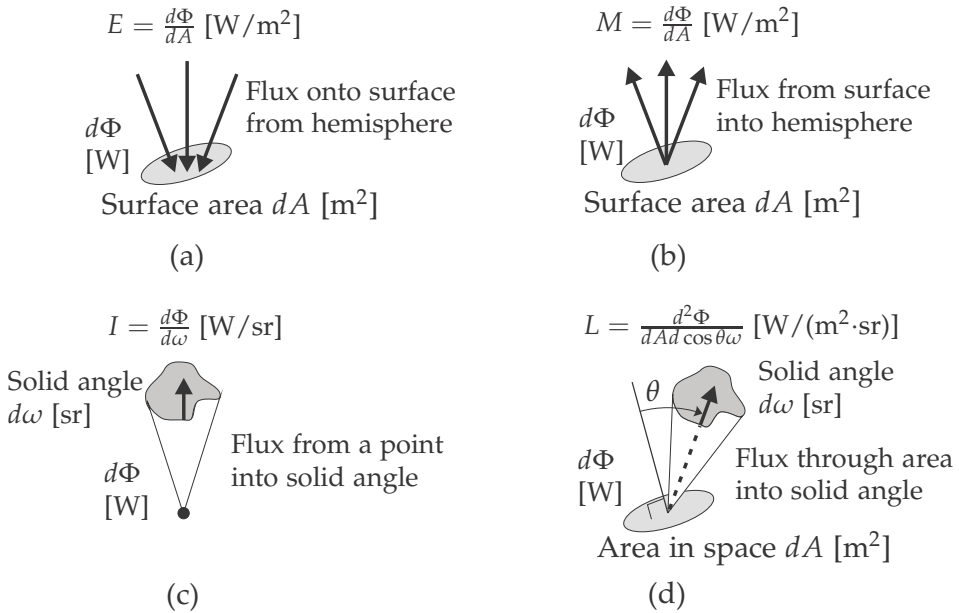


Figure 2.3 Graphic representation of radiometric quantities: (a) irradiance, (b) exitance, (c) intensity, and (d) radiance.

Table 2.1 Nomenclature for radiometric quantities.

Basic Quantity	Radiant	Photon	Photometry
Energy	Radiant energy Q_e joule [J]	Photon count Q_p, Q_q photon count [q]	Luminous energy Q_v lumen-second [lm·s]
Flux	Radiant flux Φ_e watt [W]	photon flux Φ_p, Φ_q photon rate [q/s]	Luminous flux Φ_v lumen [lm]
Exitance	Radiant exitance M_e [W/m ²]	Photon exitance M_p, M_q [q/(s·m ²)]	Luminous exitance M_v [lm/m ²]
Intensity	Radiant intensity I_e [W/sr]	Photon intensity I_p, I_q [q/(s·sr)]	Luminous intensity I_v candela=cd=[lm/sr]
Radiance	Radiance L_e [W/(m ² ·sr)]	Photon radiance L_p, L_q [q/(s·m ² ·sr)]	Luminance L_v nit=nt=[lm/(m ² ·sr)]
Irradiance	Radiant irradiance E_e [W/m ²]	Photon irradiance E_p, E_q [q/(s·m ²)]	Illuminance E_v lux=lx=[lm/m ²]
Subscript	e	p or q	v

measurement. The units are $[W/sr]$, $[q/(s \cdot sr)]$, or $[lm/sr]$, depending on the nature of the quantity.

Radiance (sterance) is the density of flux per unit source surface area and unit solid angle, as shown in Figure 2.3(d). Radiance is a property of the electromagnetic field irrespective of spatial location (in a lossless medium). For a radiating surface, the radiance may comprise transmitted light, reflected light, emitted light, or any combination thereof. The field's radiance applies anywhere in space, also on the receiving surface (see Section 2.6.1). The source must have a nonzero size. All radiance quantities are denoted by the symbol L . The human perception of 'brightness' of a large surface can be likened to a radiance experience (beware of the nonlinear response in the eye, however). The units are $[W/(m^2 \cdot sr)]$, $[q/(s \cdot m^2 \cdot sr)]$, or $[lm/(m^2 \cdot sr)]$, depending on the nature of the quantity.

When it is necessary to indicate the nature of the interaction of the flux with a physical entity, this may be done by appending subscripts. The flux may be incident Φ_i , reflected Φ_r , absorbed Φ_a , transmitted Φ_t , or emitted (exitant) Φ_m . This notation is not standard practice and should be duly documented when used.

2.3.2 Nature of radiometric quantities

The unit of the quantity is denoted by subscripting the quantity symbol by either an e (radiant energy units $[W]$), p , q (photon rate units $[q/s]$), or v (photometric units $[lm]$), as indicated in Table 2.1. If the nature of the quantity is clear from the context of usage, the subscript is not used, i.e., if all calculations are performed in photon rate quantities, no subscript is used because the dimensional units will indicate $[q/s]$. It is advisable to use the subscripts whenever the context could be misleading. In this book, if no subscripts are used, and the context is not clear, radiant units $[W]$ are assumed. Each of these quantities are briefly described in Table 2.1.

2.3.3 Spectral quantities

Three spectral domains are commonly used: wavelength λ in $[m]$, frequency ν in $[Hz]$, and wavenumber $\tilde{\nu}$ in $[cm^{-1}]$ (the number of waves that will fit into a 1-cm length). Spectral quantities indicate an amount of the quantity within a small spectral width $d\lambda$ around the value of λ : it is a spectral density. Spectral density quantity symbols are subscripted with a λ or ν , i.e., L_λ or L_ν . The dimensional units of a spectral density quantity are indicated as $[(\mu m)^{-1}]$ or $[(cm^{-1})^{-1}]$, i.e., $[W/(m^2 \cdot sr \cdot \mu m)]$.

The relationship between the wavelength and wavenumber spectral

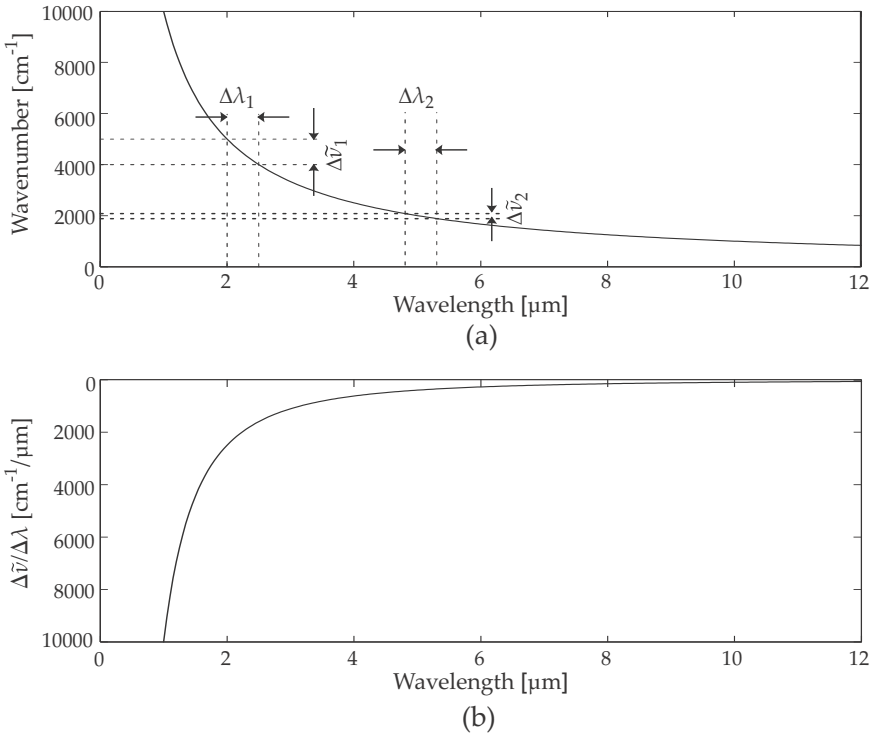


Figure 2.4 The relationship between (a) wavelength and wavenumber, and (b) $d\tilde{\nu}/d\lambda$ versus wavelength.

domains is $\tilde{\nu} = 10^4/\lambda$, where λ is in units of μm . The conversion of a spectral density quantity such as $[\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \text{cm}^{-1})]$ requires the derivative, $d\tilde{\nu} = -10^4 d\lambda/\lambda^2 = -\tilde{\nu}^2 d\lambda/10^4$. The derivative relationship converts between the spectral *widths*, and hence the spectral densities, in the two respective domains. The conversion from a wavelength spectral density quantity to a wavenumber spectral density quantity is $dL_{\tilde{\nu}} = dL_{\lambda} \lambda^2/10^4 = dL_{\lambda} 10^4/\tilde{\nu}^2$.

The relationships between wavelength, wavenumber, and the derivative $\Delta\tilde{\nu}/\Delta\lambda \approx d\tilde{\nu}/d\lambda$ are shown in Figure 2.4. The top graph clearly illustrates how a constant spectral width in the wavelength domain translates to a nonconstant spectral width in the wavenumber domain.

Spectral quantities denote the amount in a small spectral width $d\lambda$ around a wavelength λ . It follows that the total quantity over a spectral range can be determined by integration (summation) over the spectral range of interest:

$$L = \int_{\lambda_1}^{\lambda_2} L_{\lambda} d\lambda. \quad (2.1)$$

The above integral satisfies the requirements of dimensional analysis (see Chapter 10) because the units of L_λ are $[W/(m^2 \cdot sr \cdot \mu m)]$, whereas $d\lambda$ has the units of $[\mu m]$, and L has units of $[W/(m^2 \cdot sr)]$.

2.3.4 Material properties

The flux incident on an object (gas, liquid, or solid) can be reflected, absorbed, or transmitted so that¹⁹

$$\Phi_{\text{incident}} = \Phi_{\text{absorbed}} + \Phi_{\text{transmitted}} + \Phi_{\text{reflected}}, \quad (2.2)$$

and, therefore

$$1 = \alpha + \tau + \rho, \quad (2.3)$$

where α represents the absorbed fraction, τ the transmitted fraction, and ρ the reflected fraction. Material properties are described in more detail in other publications.^{4,20–22}

Some texts advocate the terminology where the suffix *-ance* is used for the characteristics of a specific sample (such as ‘transmittance of this filter’), and the suffix *-ivity* is used for the intrinsic properties of (pure) materials (such as ‘reflectivity of gold’). This terminology is not standardized²⁰ and is not strictly followed in this book.

2.4 Linear Angle

Consider the two *different* linear angles $\angle ABC$ or $\angle ABD$ in Figure 2.5. Consider the angle $\angle ABC$: project the points A and C onto the circle to obtain the points A' and C' . Likewise, project the points A and D onto the circle to obtain the points A' and D' . Hence, the *direction* from B to points A , C , and D , irrespective of distance, are projected onto a circle with radius r . From the definition of linear angle, the *equal* value of both these angles is given by $\theta = s/r$ in radians, where s is the arc length between the projected points A' and C' or D' .

The algorithm for linear angle measurement is therefore: (1) project the directions of A , C , and D as seen from B onto a circle centered on B , (2) measure the arc length between the projected points, and (3) divide the arc length by the circle’s radius.

For a full revolution, the arc length is the circumference of a circle or $2\pi r$, hence the linear angle of one revolution is 2π rad. A right angle (90 deg) has an arc length of $2\pi r/4$, leading to a linear angle of $\pi/2$ rad.

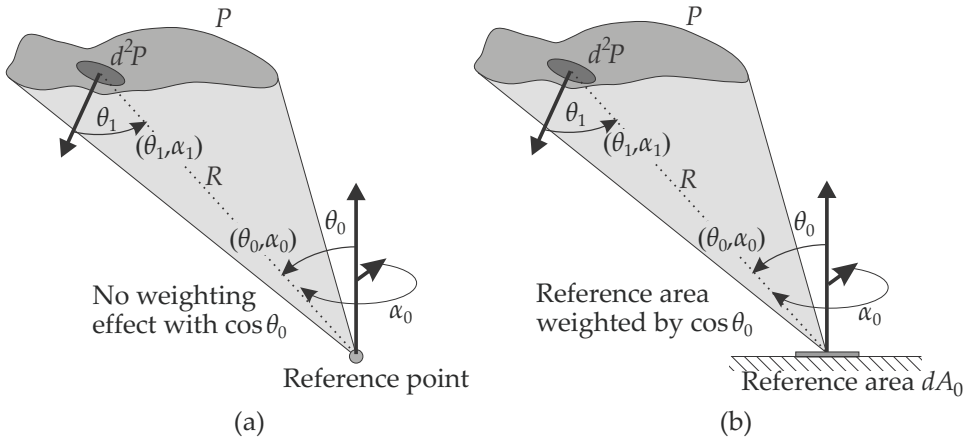


Figure 2.6 Solid angle definitions: (a) geometric solid angle, and (b) projected solid angle.

the reference point. Hence, a given area at a given distance will always have the same geometric solid angle irrespective of the direction of the area.

The *projected* solid angle Ω of any arbitrary surface P from the reference area dA_0 is given by

$$\Omega = \iint^P \frac{d^2P \cos \theta_0 \cos \theta_1}{R^2}, \quad (2.5)$$

where $d^2P \cos \theta_1$ is the projected surface area of the surface P in the direction of the reference area, and R is the distance from d^2P to the reference area. The integral depends on the viewing direction (θ_0, α_0) from the reference area, by the projected area $(dA_0 \cos \theta_0)$ of dA_0 in the direction of d^2P . Hence, a given area at a given distance will always have a different projected solid angle in different directions. The calculation of solid angles can be seen as a form of spatial normalization.

2.5.2 Geometric solid angle of a cone

Consider the left side of Figure 2.7, where a hemisphere with radius r is constructed around the origin of the coordinate system. The center of the hemisphere is located at the origin. The geometric solid angle ω , subtended at the origin, of any arbitrary surface P is given by

$$\omega = \frac{P'}{r^2}, \quad (2.6)$$

where P' is the projection of the surface P onto a sphere of radius r , as shown in Figure 2.7. The dimensional unit of the solid angle is $[\text{m}^2/\text{m}^2]$,

Geometric solid angle

Projected solid angle

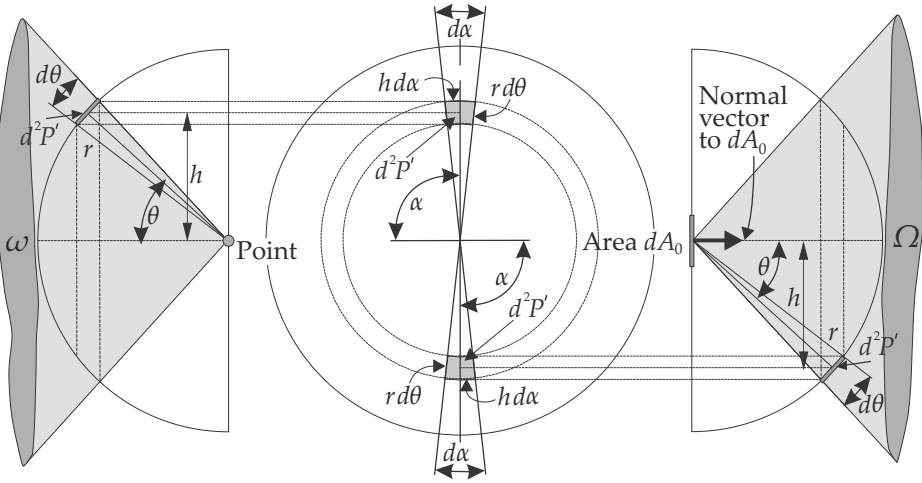


Figure 2.7 Geometric solid angle ω and projected solid angle Ω of a cone.

indicated with [sr], where the numerator is area, and the denominator is radius squared.

Note that the geometric solid angle is independent of the direction of the true area P ; only the projected area P' is relevant.

For a cone with half-apex angle of Θ (the full-apex angle is therefore 2Θ), the geometric solid angle can be derived as shown below. A small portion of the projected surface area d^2P' can be written as

$$\begin{aligned} d^2P' &= rd\theta \times h d\alpha \\ &= r^2 d\theta \sin \theta d\alpha. \end{aligned} \quad (2.7)$$

Apply Equation (2.6) and integrate over $0 \leq \theta \leq \Theta$ and $0 \leq \alpha \leq 2\pi$ rad to obtain

$$\begin{aligned} \omega &= \int_0^{2\pi} \int_0^\Theta \frac{r^2 d\theta \sin \theta d\alpha}{r^2} \\ &= 2\pi \int_0^\Theta \sin \theta d\theta \\ &= 2\pi(1 - \cos \Theta) \\ &= 4\pi \sin^2 \left(\frac{\Theta}{2} \right). \end{aligned} \quad (2.8)$$

For $\Theta = \pi/2$ rad, the cone covers a full hemisphere, and the solid angle is 2π sr; for $\Theta = \pi$ rad, the full sphere has a solid angle of 4π sr.

2.5.3 Projected solid angle of a cone

The projected solid angle is calculated relative to a small surface dA_0 located at the origin (not a single point as for the geometric solid angle). The small surface dA_0 has a vector normal to the surface, as shown on the right side of Figure 2.7.

The projected area P' is weighted with a factor $\cos \theta$, where θ is the angle between the normal vector of the small surface dA_0 and the direction to the projected area P' .

The contribution of the projected area P' to the solid angle therefore depends on the direction of the projected area P' , relative to the normal vector of the area dA_0 . Restated, the direction of the projected area P' significantly determines its contribution to the solid angle. The term 'projected solid angle' stems from the fact that the projection of the small area dA_0 at the origin weights the contribution of the projected area P' . The reason for doing this is described in Section 2.6.

The projected solid angle is calculated by considering a small area on the sphere as follows:

$$\begin{aligned} d^2P' &= rd\theta \times h d\alpha \\ &= r^2 d\theta \sin \theta d\alpha. \end{aligned} \quad (2.9)$$

Weigh the surface area with a factor $\cos \theta$ and integrate over $0 \leq \theta \leq \Theta$ and $0 \leq \alpha \leq 2\pi$ rad to obtain

$$\Omega = \int_0^{2\pi} \int_0^\Theta \frac{\cos \theta r^2 d\theta \sin \theta d\alpha}{r^2} \quad (2.10)$$

$$\begin{aligned} &= 2\pi \int_0^\Theta \cos \theta \sin \theta d\theta \\ &= \pi \sin^2 \Theta. \end{aligned} \quad (2.11)$$

Therefore, for a cone with half-apex angle of Θ , the projected solid angle is given by

$$\Omega = \pi \sin^2 \Theta. \quad (2.12)$$

For $\Theta = \pi/2$ rad, the cone covers a full hemisphere, but the projected solid angle is only π sr, one-half of the geometric solid angle. This difference is due to the fact that the area d^2P' is weighted by the cosine of θ .

For cone half-apex angles $\Theta < 0.2$ rad, the geometric solid angle and projected solid angle are numerically similar because $\cos \theta$ is near unity

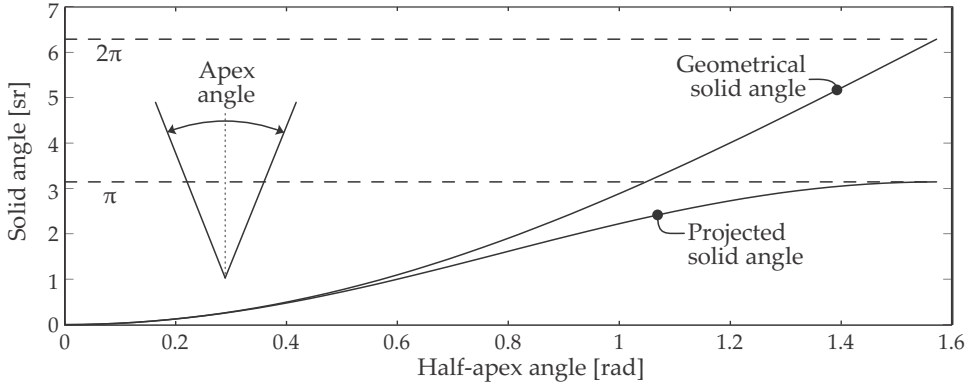


Figure 2.8 Comparison of geometrical and projected solid angle of a cone.

and the projection effect is negligible. Figure 2.8 shows the geometrical and projected solid angle magnitude for a cone.

2.5.4 Geometric solid angle of a flat rectangular surface

The geometric solid angle of a rectangular flat surface, with dimensions W and D , as seen from a reference point centered above the surface (see Figure 2.9), is determined by the integral of the projected area of a small elemental area $\cos \theta \, dd \, dw$ across the full size of the surface:

$$\begin{aligned}
 \omega_s &= \int_W \int_D \frac{dw \, dd \, \cos \theta}{R^2} \\
 &= \int_W \int_D \frac{dw \, dd \, \cos^3 \theta}{H^2} \\
 &= \int_W \int_D \frac{dw \, dd}{H^2} \left(\frac{H}{R} \right)^3 \\
 &= \int_W \int_D \frac{dw \, dd}{H^2} \left(\frac{H}{\sqrt{w^2 + d^2 + H^2}} \right)^3, \quad (2.13)
 \end{aligned}$$

where H is the reference point height above the surface. The integral is performed along the W and D dimensions with increments of dw and dd . The slant range between the reference point and the elemental area $dd \times dw$ is $R = H / \cos \theta$.

2.5.5 Projected solid angle of a flat rectangular surface

The projected solid angle of a rectangular flat surface, as seen from a reference area centered above the surface and parallel to the surface, is determined by the integral of the projected area of a small elemental area

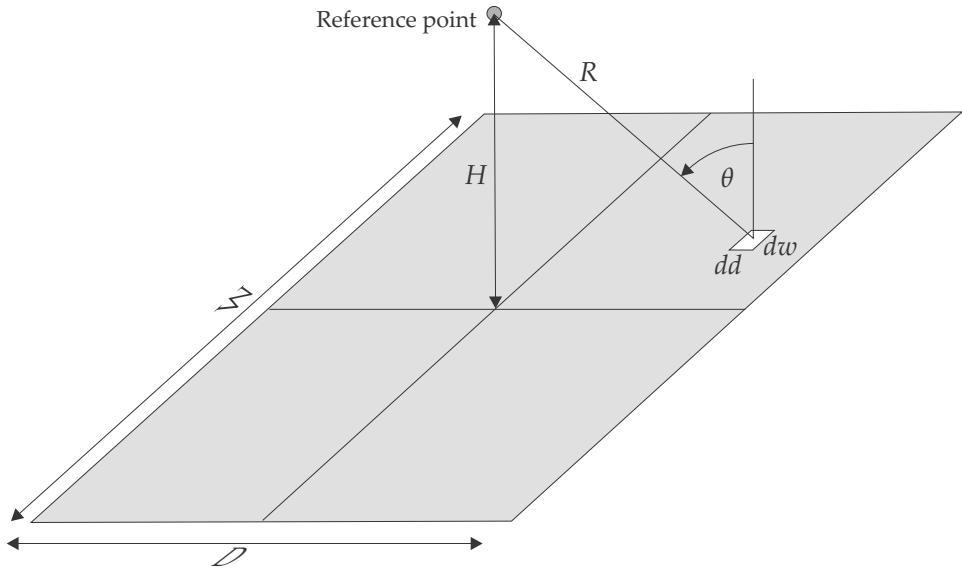


Figure 2.9 Solid angle of a centered flat surface.

$\cos \theta \, dd \, dw$, weighted by an additional $\cos \theta$ across the full size of the surface:

$$\begin{aligned}
 \Omega_s &= \int_W \int_D \frac{dw \, dd \, \cos^2 \theta}{R^2} \\
 &= \int_W \int_D \frac{dw \, dd \, \cos^4 \theta}{H^2} \\
 &= \int_W \int_D \frac{dw \, dd}{H^2} \left(\frac{H}{R} \right)^4 \\
 &= \int_W \int_D \frac{dw \, dd}{H^2} \left(\frac{H}{\sqrt{w^2 + d^2 + H^2}} \right)^4. \tag{2.14}
 \end{aligned}$$

Note the $\cos^4 \theta$ term in the integral. At large θ , i.e., when two large surfaces are located relatively close together, the contribution of the extreme areas of the surfaces reduce considerably compared to the central areas. In optical systems, the $\cos^4 \theta$ effect results in lower image flux toward large field angles — the image flux is not constant across the whole image field.

2.5.6 Approximation of solid angle

Consider a flat plate with dimensions W and D , tilted at an angle θ with respect to the view direction. It is tempting to simplify the calculation of

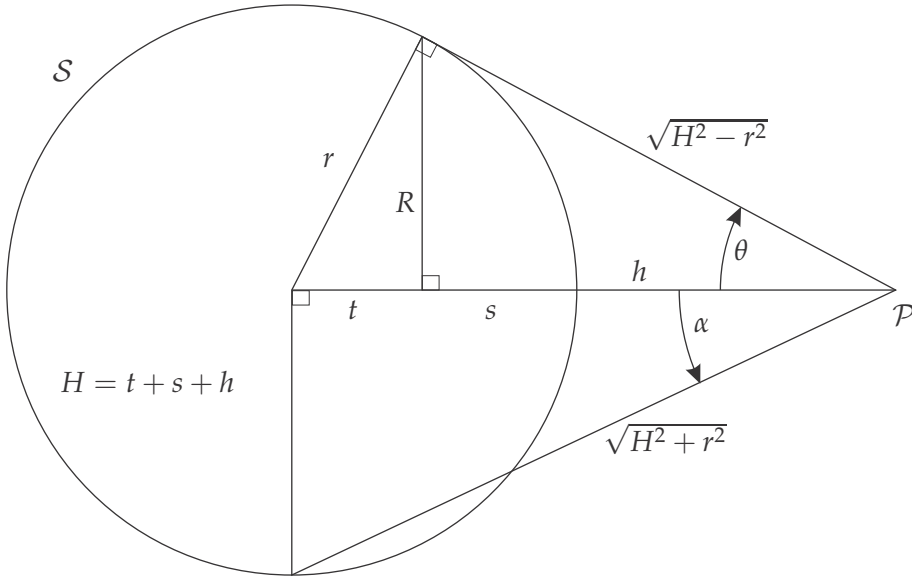


Figure 2.10 Projected area of a sphere.

the flat plate's solid angle by considering θ constant, such that

$$\begin{aligned}\omega &= \int_W \int_D \frac{\cos \theta \, dw \, dd}{R^2} \\ &= \frac{\cos \theta \, W \, D}{R^2}.\end{aligned}\tag{2.15}$$

This approximation may be acceptable only if R^2 is much larger than the projected area, such that $\cos \theta$ approaches a constant value over the full projected area. If the value of $\cos \theta$ varies appreciably across the projected area, its effect must be included in the integral, as shown in the previous section. Section E.2 investigates the accuracy of the approximation.

2.5.7 Projected area of a sphere

It is commonly assumed that the projected area of a sphere is a circle with the same diameter as the sphere. This is only true when the sphere is observed from infinity. At any closer range, the projected area of the sphere is determined by how much of the sphere is visible. Figure 2.10 shows that the projected area at close range is given by πR^2 , whereas the area of a disk is given by πr^2 , where $R < r$ near the sphere. The solid angle of the sphere at close range can indeed be larger than the solid angle calculated from a disk at the center of the sphere, as shown in the next section.

2.5.8 Projected solid angle of a sphere

When observing sphere \mathcal{S} from reference point \mathcal{P} in Figure 2.10, only a portion of the sphere is visible. With \mathcal{P} closer to \mathcal{S} , less of the sphere is visible. As the distance between \mathcal{P} and \mathcal{S} approaches infinity, a full hemisphere is observed.

The projected solid angle of a cone is given by $\Omega = \pi \sin^2 \theta$, where θ is the half-apex angle. From Figure 2.10 it is evident that for the sphere, the solid angle of the cone defined by the visible portion of the sphere (with half-apex angle θ) is given by

$$\Omega_s = \pi \sin^2(\arcsin r/H) = \pi r^2/H^2. \quad (2.16)$$

The solid angle of the cone, defined by the disk at the center of the sphere with normal vector pointed to \mathcal{P} and with half-apex angle α , is given by

$$\Omega_d = \pi \sin^2(\arcsin r/\sqrt{H^2 + r^2}) = \pi r^2/(H^2 + r^2). \quad (2.17)$$

The ratio of sphere solid angle to disk solid angle is then

$$\frac{\Omega_s}{\Omega_d} = \frac{r^2 + H^2}{H^2}. \quad (2.18)$$

This ratio is always greater than one, hence the solid angle when viewing a sphere from a distance is always greater than the solid angle of a disk at the center of the sphere (both referenced from \mathcal{P}). For increasing H the ratio approaches unity, but at close range the ratio increases.

For the smallest value of H ($H = r$) the ratio is two. At this location, the surface of the sphere, the projected solid angle of the sphere is π because the sphere appears locally as an infinitely large, almost-planar surface. At the same location, the disk-projected solid angle is $\pi \sin^2(\pi/4) = \pi/2$ sr.

2.6 Radiance and Flux Transfer

2.6.1 Conservation of radiance

The conservation of radiance⁹ is an important, fundamental concept in the understanding of radiometry. In this section, the ‘small area’ (derivative) notation is used for flux, area, and solid angle, with the understanding that it applies to small elemental-source and receiver areas. The general solution for arbitrary source and receiver areas is addressed in Section 6.6.

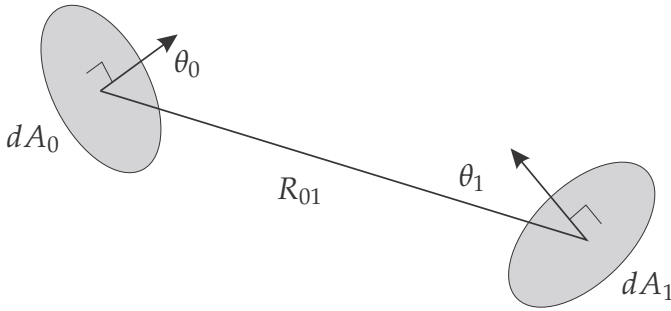


Figure 2.11 Geometrical construction for radiative flux between two elemental areas.

Radiance is conserved for flux propagation through a lossless optical medium.⁷ Consider the construction in Figure 2.11: two elemental areas dA_0 and dA_1 are separated by a distance R_{01} , with the angles between the normal vector of each surface and the line of sight given by θ_0 and θ_1 . A total flux of Φ is flowing through both the surfaces. Then, from the definition of radiance in Section 2.3, radiance values at both surfaces are

$$L_0 = \frac{d^2\Phi}{dA_0 \cos \theta_0 d\Omega_1} \quad (2.19)$$

and

$$L_1 = \frac{d^2\Phi}{dA_1 \cos \theta_1 d\Omega_0}. \quad (2.20)$$

From the definition of solid angle,

$$d\Omega_1 = \frac{dA_1 \cos \theta_1}{R_{01}^2} \quad (2.21)$$

and

$$d\Omega_0 = \frac{dA_0 \cos \theta_0}{R_{01}^2}. \quad (2.22)$$

The same flux flows through both surfaces, and hence the flux in Equations (2.19) and (2.20) is the same. After mathematical manipulation, it follows that $L_0 = L_1$. It is important to note that there are no restrictions on the location of either dA_0 or dA_1 ; it follows that radiance is spatially invariant *in any plane in the system* (provided that the field is not affected by an object or medium losses). Radiance is a property of the electromagnetic field itself; it will be affected by geometrical constructs in its surroundings but is not dependent on such constructs.

As light propagates through mediums with different refractive indices n such as air, water, glass, etc., the entity called *basic radiance*,^{6,7,9} defined by L/n^2 , is invariant. It can be shown that for light propagating from a medium with refractive index n_1 to a medium with refractive index n_2 , the basic radiance is conserved:

$$\frac{L_1}{n_1^2} = \frac{L_2}{n_2^2}. \quad (2.23)$$

Most mediums are not lossless; a medium may attenuate the propagating flux by removing flux from the beam through scattering or absorption, or the medium may add flux to the beam through scattering or thermal exitance. The effects of a lossy medium are discussed in more detail in Section 4.2.

2.6.2 Flux transfer through a lossless medium

A lossless medium is defined as a medium with no losses between the source and the receiver, such as a complete vacuum. This implies that no absorption, scattering, or any other attenuating mechanism is present in the medium.

Combining Equations (2.19) and (2.21) yields

$$L = \frac{d^2\Phi R_{01}^2}{dA_0 \cos\theta_0 dA_1 \cos\theta_1}, \quad (2.24)$$

or

$$d^2\Phi = \frac{L dA_0 \cos\theta_0 dA_1 \cos\theta_1}{R_{01}^2}. \quad (2.25)$$

For a lossless medium, the flux flowing between the source and receiver is given by the product of the (invariant) radiance and the projected areas of the source and receiver, divided by the square of the distance between the areas. Note that on the right of the equation, there is only one radiometric quantity, L ; the remaining quantities are all geometric (nonradiometric) quantities. Radiometry is therefore as much a study of geometry as it is of optical flux. Equation (2.25) pertains to the flux flowing through the two surfaces; it does not yet include the effects of source emissivity or receiver absorption (see Section 3.2).

Equation (2.25) can be used to derive all radiometric quantities as follows:

Irradiance is derived as (note the $\cos\theta_1$ term)

$$E = \frac{d\Phi}{dA_1} = \frac{L dA_0 \cos\theta_0 \cos\theta_1}{R_{01}^2} = L \Omega_0 \cos\theta_1. \quad (2.26)$$

Intensity is derived as

$$I = \frac{d\Phi R_{01}^2}{dA_1} = L dA_0 \cos \theta_0 \cos \theta_1. \quad (2.27)$$

Exitance is derived as (note the $\cos \theta_0$ term)

$$M = \frac{d\Phi}{dA_0} = \frac{L dA_1 \cos \theta_1 \cos \theta_0}{R_{01}^2} = L \Omega_1 \cos \theta_0. \quad (2.28)$$

A Lambertian source radiates into the full hemisphere with projected solid angle π , so that, for Lambertian sources, Equation (2.28) reduces to $M = L \pi$ (see Section 2.7).

See Equation (10.6) for a discussion on manipulating the dimensional units of these equations.

2.6.3 Flux transfer through a lossy medium

Denote the medium transmittance by τ_{01} to indicate that it is the transmittance between location 0 and location 1. The total flux flowing from the source *through* the receiver area is given by

$$d^2\Phi_1 = \frac{L_0 dA_0 \cos \theta_0 dA_1 \cos \theta_1 \tau_{01}}{R_{01}^2}, \quad (2.29)$$

where Φ_1 is the flux at dA_1 , L_0 is the radiance at dA_0 , and τ_{01} is the medium transmittance between the two elemental areas. In this case radiance is *not* invariant because of the medium loss. This simple model is accurate for most cases where the path radiance contribution is negligible compared to the flux in the field radiance.

2.6.4 Sources and receivers of arbitrary shape

Equation (2.25) calculates the flux flowing between two infinitely small areas. The flux flowing between two arbitrary shapes can be calculated by integrating Equation (2.25) over the source surface and the receiving surface. In the general case, the radiance L cannot be assumed constant over A_0 , introducing the spatial radiance distribution $L(dA_0)$ as a factor into the spatial integral.⁹ Likewise, the medium transmittance between any two areas dA_0 and dA_1 varies with the spatial locations of dA_0 and dA_1 — hence $\tau_{01}(dA_0, dA_1)$ should also be included in the spatial integral.

The integral can be performed over any arbitrary shape, as shown in Figure 2.12, supporting the solution to complex geometries. Clearly

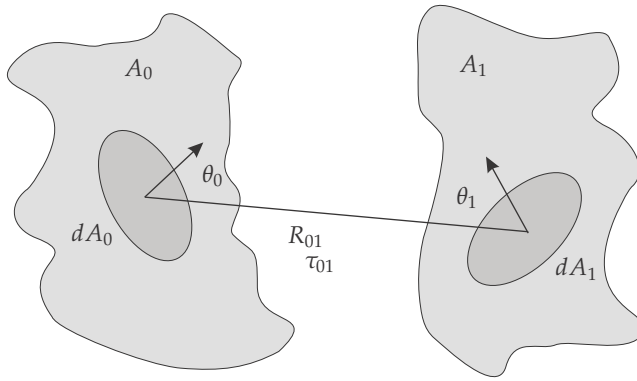


Figure 2.12 Radiative flux between areas of arbitrary shape.

matters such as obscuration and occlusion should be considered when performing this integral:

$$\Phi = \int_{A_0} \int_{A_1} \frac{L(dA_0) dA_0 \cos \theta_0 dA_1 \cos \theta_1 \tau_{01}(dA_0, dA_1)}{R_{01}^2}. \quad (2.30)$$

2.6.5 Multi-spectral flux transfer

The optical power leaving a source undergoes a succession of scaling or ‘spectral filtering’ processes as the flux propagates through the system, as shown in Figure 2.13. This filtering varies with wavelength. Examples of such filters are source emissivity, atmospheric transmittance, optical filter transmittance, and detector responsivity. The multi-spectral filter approach described here is conceptually simple but fundamental to the calculation of radiometric flux. Consider the flow of flux from the source to the sensor:

1. The most fundamental description of a thermal radiator source state is given by the temperature of the radiating surface and a mathematical function, called Planck’s law. At a given temperature, Planck’s law sets an absolute limit to the source radiance (see Section 3.1).
2. The spectral emissivity of the source acts as a filter by limiting the source radiance. Emissivity can be expressed as a spectral variable between zero and unity (see Section 3.2).
3. The spectral transmittance of the medium or atmosphere acts as a spectral filter (see Chapter 4).

Multi-spectral flux propagating from the source to the sensor.
Calculate flux at a single wavelength, then add the resultant flux across all wavelengths.

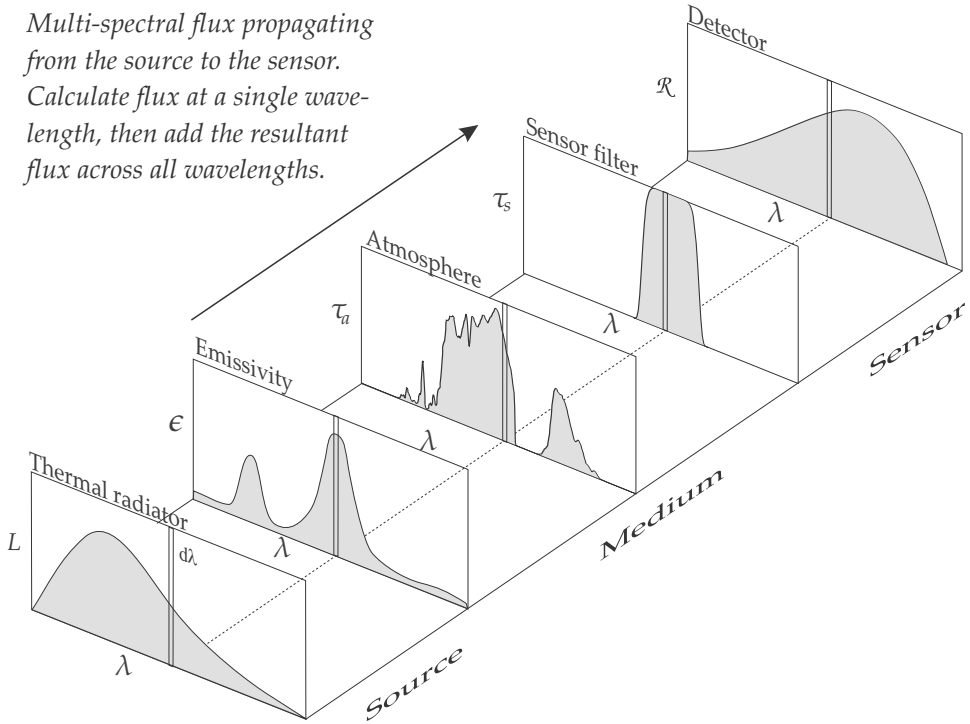


Figure 2.13 Describing the electro-optical system as a thermal source and a series of spectral filters.

4. In some cases the source radiance is reflected from a surface, such as sunlight reflected from the surface of an object. The spectral nature of the reflectance can be considered as a spectral filter (see Section 3.4).
5. The spectral transmittance of the optics/filter in the sensor acts as a filter (see Section 6.4).
6. The detector's spectral response can be interpreted as a spectral filter (see Chapter 5).
7. The detector converts the optical flux to an electrical signal by the scalar value of its responsivity (see Chapter 5).

Extend Equation (2.29) for multi-spectral calculations by noting that over a spectral width $d\lambda$ the radiance is given by $L = L_\lambda d\lambda$:

$$d^3\Phi_\lambda = \frac{L_{0\lambda} dA_0 \cos \theta_0 dA_1 \cos \theta_1 \tau_{01} d\lambda}{R_{01}^2}, \quad (2.31)$$

where $d^3\Phi_\lambda$ is the total flux in [W] or [q/s] flowing in a spectral width $d\lambda$ at wavelength λ , from a radiator with radiance $L_{0\lambda}$ with units [W/(m²·sr·μm)]

and projected surface area $dA_0 \cos \theta_0$, through a receiver with projected surface area $dA_1 \cos \theta_1$ at a distance R_{01} , with a transmittance of τ_{01} between the two surfaces. The transmittance τ_{01} now includes all of the spectral variables in the path between the source and the receiver.

To determine the *total flux* flowing from elemental area dA_0 through dA_1 over a wide spectral width, divide the wide spectral band into a large number N of narrow widths $\Delta\lambda$ at wavelengths λ_n and add the flux for all of these narrow bandwidths together as follows:

$$d^2\Phi = \sum_{n=0}^N \left(\frac{L_{0\lambda_n} dA_0 \cos \theta_0 dA_1 \cos \theta_1 \tau_{01\lambda_n} \Delta\lambda}{R_{01}^2} \right). \quad (2.32)$$

By the Riemann–Stieltjes theorem in reverse, if now $\Delta\lambda \rightarrow 0$ and $N \rightarrow \infty$, the summation becomes the integral

$$d^2\Phi = \int_{\lambda_1}^{\lambda_2} \frac{L_{0\lambda} dA_0 \cos \theta_0 dA_1 \cos \theta_1 \tau_{01\lambda} d\lambda}{R_{01}^2}. \quad (2.33)$$

Equation (2.33) describes the total flux at all wavelengths in the spectral range λ_1 to λ_2 passing through the system. This equation is developed further in Chapter 7.

2.7 Lambertian Radiators and the Projected Solid Angle

“A Lambertian source is, by definition, one whose *radiance* is completely independent of viewing angle.”⁷ Many (but not all) rough and natural surfaces produce radiation whose radiance is approximately independent of the angle of observation. These surfaces generally have a rough texture at microscopic scales. Planck-law blackbody radiators are also Lambertian sources (see Chapter 3.1). Any Lambertian radiator is completely described by its scalar radiance magnitude only, with no angular dependence in radiance.

The relationship between the exitance and radiance for such a Lambertian surface can be easily derived. If the flux radiated from a Lambertian surface Φ [W] is known, it is a simple matter to calculate the exitance $M = \Phi/A$ [W/m²], where A is the radiating surface area. Because radiance has units [W/(m²·sr)], it may appear tempting to divide the exitance by the total hemispherical solid angle to determine the radiance.

The correct relationship between exitance and radiance for a Lambertian source is derived by starting with Equation (2.25):

$$d^2\Phi = \frac{L dA_0 \cos \theta_0 dA_1 \cos \theta_1}{R_{01}^2}. \quad (2.34)$$

This equation describes the flux flowing between two small areas. The source radiates power in all directions, and in order to determine the power radiated from the source into the hemisphere, one must integrate the receiver area $dA_1 \cos \theta_1$ over the whole hemisphere. In order to perform the integration, construct a hemispherical dome with radius r and its center at the elemental source area dA_0 , as shown in Figure 2.7. The projected receiver area can be written as

$$dA_1 \cos \theta_1 = r d\theta_0 \times h d\alpha. \quad (2.35)$$

Assume a constant radiance over the small elemental source area and integrate over the complete hemisphere to obtain

$$\Phi_H = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \frac{L dA_0 \cos \theta_0 r^2 d\theta_0 \sin \theta_0 d\alpha}{r^2} \quad (2.36)$$

$$\begin{aligned} &= LA_0 \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \cos \theta_0 \sin \theta_0 d\theta_0 d\alpha \\ &= LA_0 2\pi \int_0^{\frac{\pi}{2}} \cos \theta_0 \sin \theta_0 d\theta_0 \\ &= LA_0 \pi, \end{aligned} \quad (2.37)$$

then

$$M = \frac{\Phi}{A_0} = L\pi. \quad (2.38)$$

This result indicates that the exitance of a Lambertian radiator is related to radiance by the projected solid angle [Equation (2.12)] of π sr, *not* the geometric solid angle of 2π sr. Why? In Equation (2.36) there is a $\cos \theta_0$ term describing the projected area of the source. A *flat* Lambertian source therefore radiates with a cosine distribution *and not isotropically in all directions*. The projected solid angle is effectively calculated by weighting the projected area of the source with the $\cos \theta_0$ term; see Section 2.5.3.

The above derivation indicates that one should always use the projected solid angle instead of the geometric solid angle when dealing with Lambertian sources. In the event that the solid angle under consideration is less than the full hemisphere, use the equation presented in Section 2.5.3. Rather than memorizing rules, it is better to perform the calculation from first principles — it is easier to remember Equation (2.25) than to remember a multitude of rules.

Consider a Lambertian source as shown in Figure 2.14(a). As the area dA is rotated, the projected area along the line to the observer decreases as $dA \cos \theta$. For a small source with area dA , when viewed at an arbitrary angle θ , the intensity varies as $I = L dA \cos \theta$. However, in Figure 2.14(b),

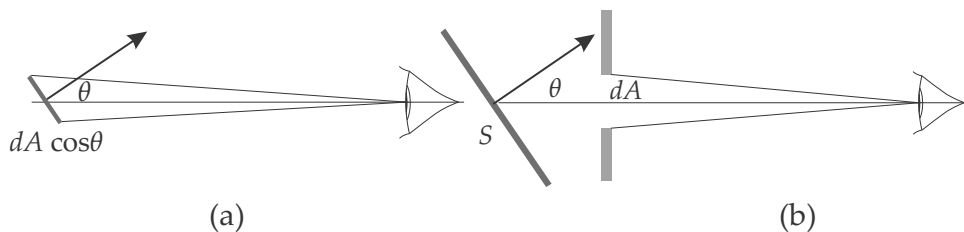


Figure 2.14 Lambertian sources: (a) finite-size tilted surface, and (b) infinite-size tilted surface behind a finite size aperture.

the area S is infinitely large compared with the window opening dA . In this case the observer always sees an area dA , independent of the angle by which the source is rotated. In the case of Figure 2.14(b), the observed intensity is given by $I = L dA$.

2.8 Spatial View Factor or Configuration Factor

It is evident from Equation (2.25) that the amount of flux transfer between two surfaces dA_0 and dA_1 has a geometrical term and a radiometric term. Equation (2.30) shows that the geometry can be calculated over areas A_0 and A_1 . A flux transfer calculation is therefore as much a geometrical calculation as it is a radiance calculation. On condition that (1) the source spatial radiance is uniform, (2) there are no medium losses, and (3) the receiving spatial area is uniform, the radiance term can be mathematically separated from the purely geometrical term. The radiometric term is a function of the field only (irrespective of space), whereas the geometric term is spatial geometry only (irrespective of radiance field considerations). The calculation of view factors can be seen as a form of spatial normalization.

The efficient transfer of heat is of prime importance when designing furnaces. The heat-transfer community developed a detailed mathematical concept for spatial geometric integrals, called the configuration factor, view factor, diffuse shape factor, or angle factor.²³ Tables of configuration factor values are precalculated for given geometrical configurations.

Assuming diffuse Lambertian surfaces for A_0 and A_1 , the view factor is the portion of all of the flux (in the hemisphere, h_s) leaving A_0 that passes through A_1 and is given by

$$F_{A_0 \rightarrow A_1} = \frac{\int_{A_0} \int_{A_1} L_0 \cos \theta_0 dA_0 \cos \theta_1 dA_1 / R_{01}^2}{\int_{A_0} L_0 dA_0 (\int_{h_s} \cos \theta_0 \cos \theta_1 dA_1 / R_{01}^2)}$$

$$= \frac{\int_{A_0} \int_{A_1} L_0 \cos \theta_0 dA_0 \cos \theta_1 dA_1 / R_{01}^2}{\pi \int_{A_0} L_0 dA_0}, \quad (2.39)$$

where $\int_{\text{hs}} \cos \theta_0 \cos \theta_1 dA_1 / R_{01}^2 = \pi$ sr is the solid angle integral yielding the projected solid angle of the hemisphere centered at dA_0 . The view factor has units of [sr/sr]. In Equation (2.39) the radiance field L is allowed to vary across the area A_0 , but if it is constant $L(A_0) = c$, it follows that the view factor for two small areas dA_0 and dA_1 is given by²³

$$F_{dA_0 \rightarrow dA_1} = \frac{\cos \theta_0 \cos \theta_1 dA_1}{\pi R_{01}^2}. \quad (2.40)$$

With the view factor known (probably from precalculated tables), the flux transfer from surface A_0 to A_1 can be determined by

$$\Phi_{A_0 \rightarrow A_1} = \pi L_0 A_0 F_{A_0 \rightarrow A_1}. \quad (2.41)$$

2.9 Shape of the Radiator

This section shows that the three-dimensional surface profile of a Lambertian radiator does not affect the ratio of radiance to exitance. In this case only a disk and a sphere is considered, but in principle, two radiators emitting the same flux but with different three-dimensional profiles will have the same ratio provided that the projected areas of the two shapes are the same.

2.9.1 A disk

Consider a disk uniformly radiating a total flux of Φ into the hemisphere. From the definition of radiance and exitance, it follows that

$$L = \frac{\Phi}{\iint dA_0 \cos \theta_0 d\Omega_1} \quad (2.42)$$

and

$$M = \frac{\Phi}{\int dA_0}. \quad (2.43)$$

Hence,

$$\begin{aligned} \frac{M}{L} &= \frac{\iint dA_0 \cos \theta_0 d\Omega_1}{\int dA_0} \\ &= \frac{A_0 \int \cos \theta_0 d\Omega_1}{A_0} \\ &= \int \cos \theta_0 d\Omega_1, \end{aligned} \quad (2.44)$$

where θ_0 is the angle of $d\Omega_1$ with respect to the disk normal vector. In the construction of Figure 2.7, $d\Omega_1 = d^2P'/r^2 = d\theta \sin\theta d\alpha$. Integrate over $0 \leq \theta \leq \pi/2$ sr and $0 \leq \alpha \leq 2\pi$ sr to obtain

$$\begin{aligned}\frac{M}{L} &= \int_0^{2\pi} \int_0^{\pi/2} \cos\theta d\theta \sin\theta d\alpha \\ &= 2\pi \int_0^{\pi/2} \sin\theta \cos\theta d\theta \\ &= \pi \sin^2(\pi/2) = \pi,\end{aligned}\tag{2.45}$$

which is just another way to derive Equation (2.38).

2.9.2 A sphere

Consider a sphere with radius r uniformly radiating a total flux of Φ into a full sphere. This amount of flux is equally divided between the visible side of the sphere and the 'dark side' of the sphere. The radiance must be calculated only from the visible side of the sphere. Again using the definitions of radiance and exitance, it follows that

$$\begin{aligned}\frac{M}{L} &= \frac{\int dA_0 \cos\theta_0 \int d\Omega_1}{\int dA_0} \\ &= \frac{2\pi r^2 \int d\Omega_1}{4\pi r^2} \\ &= \frac{1}{2} \int d\Omega_1.\end{aligned}\tag{2.46}$$

In the construction of Figure 2.7, $d\Omega_1 = d^2P'/r^2 = d\theta \sin\theta d\alpha$. Integrate over $0 \leq \theta \leq \pi/2$ sr (only a hemisphere is visible to contribute to radiance) and $0 \leq \alpha \leq 2\pi$ sr to obtain

$$\begin{aligned}\frac{M}{L} &= \frac{1}{2} \int_0^{2\pi} \int_0^{\pi/2} d\theta \sin\theta d\alpha \\ &= \pi \int_0^{\pi/2} \sin\theta d\theta \\ &= 2\pi \sin^2(\pi/4) = \pi.\end{aligned}\tag{2.47}$$

Considering that at any moment only half of a sphere is visible, this agrees with the ratio of radiance to exitance derived in the previous section.

2.10 Photometry and Color

2.10.1 Photometry units

Photometric quantities are not specified in spectral variables (e.g., Planck's law) but rather as normalized variables in terms of a standard source and

a standard eye spectral response. The standard source is a blackbody radiator at a temperature of 2042 K (the solidifying point of platinum), with a luminance of 6×10^5 candela/m². The standard eye response is described in more detail in the next section. Luminous intensity is expressed in candela.

2.10.2 Eye spectral response

The normalized spectral response of the eye is called the *relative luminous efficiency*. The exact shape of the relative luminous efficiency depends on the light level. The two extremes of relative luminous efficiency are known as photopic (high luminance levels) and scotopic (low luminance levels) luminous efficiencies. Unless otherwise specified, photometric values are normally specified for the photopic spectral response.

If the luminance exceeds about 3 lm/(m²·sr), the eye is light-adapted, and the cones in the retina are operating. Under light-adapted conditions, the eye's spectral response is called photopic. In photopic vision, the eye has color discrimination and acute foveal vision. The standard photopic luminous efficiency is shown in Figure 2.15 and Table A.4. The spectral shape of photopic luminous efficiency is defined in tabular form^{24,25} but can be *roughly approximated* by

$$V_{\lambda} = 1.019 \exp(-285.51(\lambda - 0.5591)^2). \quad (2.48)$$

If the luminance is less than 3×10^{-5} lm/(m²·sr), the eye is dark-adapted. The cones are no longer sensitive, and the rods sense the light. Under dark-adapted conditions, the eye's spectral response is called scotopic. Under scotopic vision, the eye is not sensitive to color and has no foveal vision. The standard scotopic luminous efficiency is shown in Figure 2.15 and Table A.4. The spectral shape of scotopic luminous efficiency is defined in tabular form^{24,25} but can be *roughly approximated* by

$$V'_{\lambda} = 0.99234 \exp(-321.1(\lambda - 0.5028)^2). \quad (2.49)$$

Equations (2.48) and (2.49) are not accurate at the extreme wavelength limits of the spectral bands. These approximations should be used with care if the source has significant amounts of flux at the wavelength limits. An example of one such case is the eye viewing a blackbody at low temperatures.

At luminance levels between 3×10^{-5} lm/(m²·sr) and 3 lm/(m²·sr), the spectral response is somewhere between the photopic and scotopic, referred to as mesopic vision.

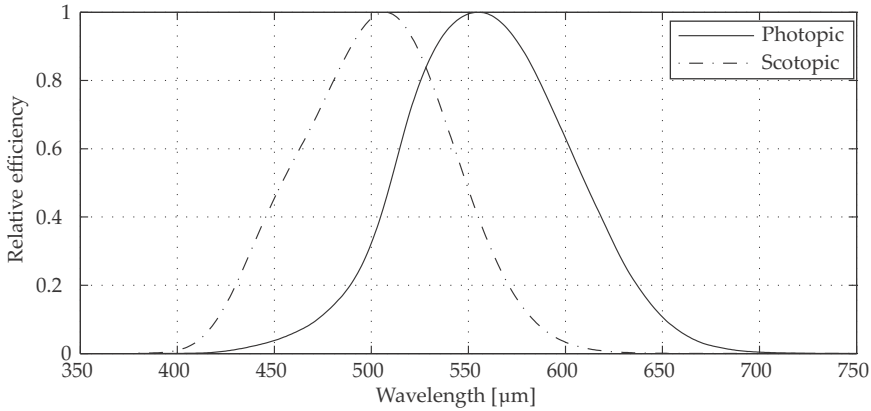


Figure 2.15 Relative luminous efficiency V_λ of the human eye.

2.10.3 Conversion to photometric units

The conversion from radiometric units (watts) to photometric units (lumens) is easily performed if the flux's spectral properties are known. This section considers radiance and luminance, but the reader should note that any radiometric quantity can be used, e.g., flux, or irradiance.

The photopic luminance of a source is defined as

$$L_v = \int_0^\infty K_{\max} V_\lambda L_{e\lambda} d\lambda, \quad (2.50)$$

where $K_\lambda = K_{\max} V_\lambda$ is the spectral photopic efficacy, V_λ is the photopic efficiency, $K_{\max} = 683 \text{ lm/W}$ is the maximum value of photopic efficacy, referenced to a 2042-K blackbody standard source, and $L_{e\lambda}$ is the source's radiance.

Likewise, the scotopic luminance of a source is defined as

$$L_v = \int_0^\infty K'_{\max} V'_\lambda L_{e\lambda} d\lambda, \quad (2.51)$$

where $K'_\lambda = K'_{\max} V'_\lambda$ is the spectral scotopic efficacy, V'_λ is the scotopic efficiency, $K'_{\max} = 1700 \text{ lm/W}$ is the maximum value of scotopic efficacy referenced to a 2042-K blackbody standard source, and $L_{e\lambda}$ is the source's radiance.

Literature sometimes shows different values for K_{\max} and K'_{\max} because the values changed every time a new standard source was instituted or a modification was made to the efficiency curves.

Efficacy can be defined as a spectral variable, as in the previous equations, or over a wide spectral band. The total (wideband) luminous efficacy

is given by the ratio of total luminance to total radiance:

$$K = \frac{L_v}{L_e} \quad (2.52)$$

$$= \frac{\int_0^\infty K_{\max} V_\lambda L_{e\lambda} d\lambda}{\int_0^\infty L_{e\lambda} d\lambda} \quad (2.53)$$

$$= \frac{K_{\max} \int_0^\infty V_\lambda L_{e\lambda} d\lambda}{\int_0^\infty L_{e\lambda} d\lambda}. \quad (2.54)$$

2.10.4 Brief introduction to color coordinates

There is a multitude of color-space definitions, each optimized for different applications. The single example provided here only serves to illustrate the key concept that the calculation of color coordinates is essentially a radiometric calculation involving normalization with given spectral weights. See Section 7.2 for more examples of normalization.

One common color space is the CIE 1931 tristimulus values XYZ, or the xyY chromaticity color space.^{24,26–28} To calculate the xyY color coordinates of a color given the spectrum, proceed as follows:

$$X(T) = \int_0^\infty \bar{x}_\lambda L_\lambda d\lambda, \quad (2.55)$$

$$Y(T) = \int_0^\infty \bar{y}_\lambda L_\lambda d\lambda, \text{ and} \quad (2.56)$$

$$Z(T) = \int_0^\infty \bar{z}_\lambda L_\lambda d\lambda, \quad (2.57)$$

where \bar{x}_λ , \bar{y}_λ , and \bar{z}_λ are the color-matching functions of the CIE standard colorimetric observer, as shown in Figure 2.16. The xyz chromaticity color coordinates can then be calculated by

$$x = \frac{X}{X + Y + Z}, \quad (2.58)$$

$$y = \frac{Y}{X + Y + Z}, \text{ and} \quad (2.59)$$

$$z = \frac{Z}{X + Y + Z} = 1 - x - y, \quad (2.60)$$

where x and y define the color coordinates in the xy chromaticity diagram (Figure A.1). Valid color coordinates are all inside the closed curve. The U- or dome-shaped part of the perimeter describes the monochromatic rainbow colors, calculated from the above equations with $L_\lambda = 1$. The color of a Planck radiator (see Section 3.1) can be calculated from the above

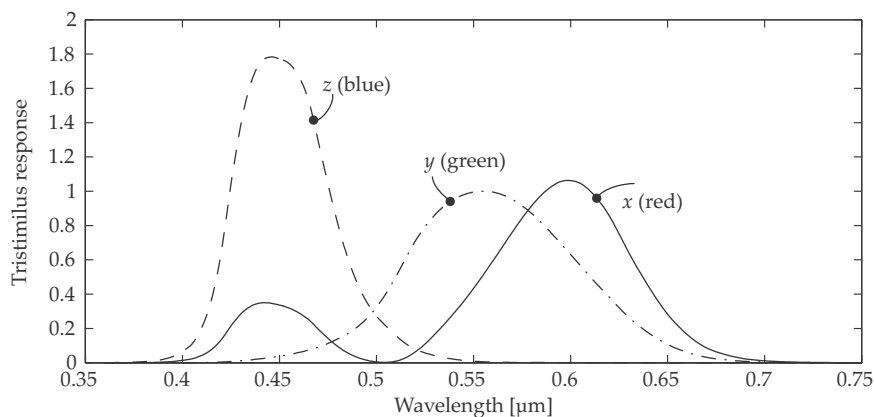


Figure 2.16 CIE standard observer color-matching functions.

equations and Equation (3.1), with $L_\lambda = M_{e\lambda}(T)/\pi$. This is known as the Planckian locus.²⁷ The coordinates of a few selected monochromatic wavelengths and the Planckian locus are shown in Figure A.1.

The intrinsic color of a surface is not necessarily the color observed in reflected light; the color coordinates of reflected light are affected by the spectral illumination onto the surface. Although there may be a surface with a perfect white color (reflecting all light equally), there is no perfect white illuminating source.²⁹ Hence, for different illuminating sources, the color white has many different color coordinate representations, known as ‘white points’.^{30,31} It was noted²⁹ that Planck radiators appear almost equally white at all temperatures above 2000 K.

2.10.5 Color-coordinate sensitivity to source spectrum

Color is an elusive property — different people perceive color differently, and the apparent color of an object depends on the illuminance spectrum. This section explores these subtleties as an application of normalizing and radiometric concepts rather than the human perception of color. The data for this analysis are available in the `pyradi`²⁵ data set.

Four sources are considered: the first light source is a ‘daylight’ fluorescent light source, the second source is the sun modeled as a thermal radiator at 5900 K, the third source is an incandescent light globe at a temperature of 2850 K, and the fourth source is a low-pressure sodium lamp. The sources’ normalized radiances are shown in Figure 2.17.

The samples illuminated by the sources are a red tomato, lettuce, a yellow prune, a dark-green leaf, a blue Nitrile (latex-like) surgical glove, and standard white printing paper. Figure 2.18 shows the spectral reflectance

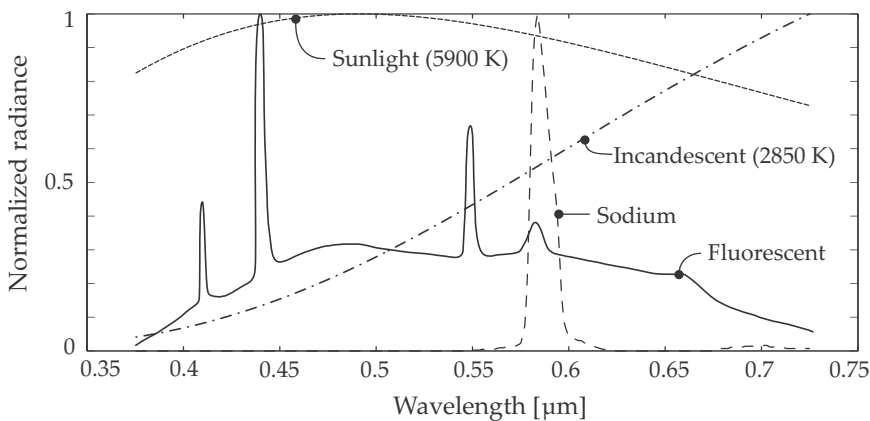


Figure 2.17 Normalized source radiance.

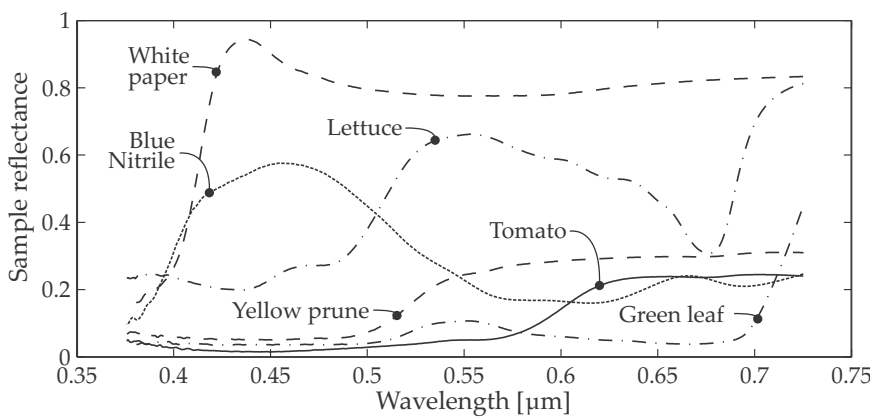


Figure 2.18 Reflectance of selected objects in the visual band.

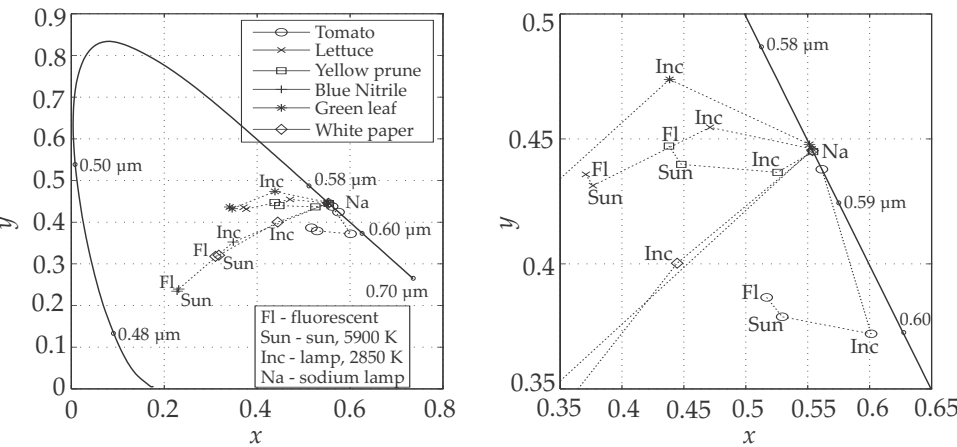


Figure 2.19 Sample color coordinates under different illumination sources.

of the samples. These diffuse reflection spectra were measured with a spectroradiometer, illuminating the sample with a bright light at short distance. The fruit samples all demonstrated considerable light propagation deeper into the fruit. The blue glove was located on top of a Spectralon white reference (note the considerable 'white' reflectance beyond 0.55 μm).

The color coordinates of the samples, when illuminated with the various sources, were calculated using Equations (2.57) and (2.60), using the code in Section D.5.4. The results are shown in Figure 2.19. It is evident that the fluorescent source is a reasonable match to daylight illumination; however, the spectral peaks in the fluorescent radiance does result in small shifts of the color coordinates relative to the sun as reference. The samples' color coordinates shifted considerably toward the orange–yellow under incandescent illumination. Finally, when illuminated by the sodium lamp, all samples had virtually the same color, 589 nm — the wavelength of the near-monochromatic illumination.

Bibliography

- [1] Wikipedia, "Electromagnetic radiation," http://en.wikipedia.org/wiki/Electromagnetic_radiation.
- [2] Wikipedia, "Photon," <http://en.wikipedia.org/wiki/Photon>.
- [3] Casagrande, M., "Birth of a Photon," <http://casagrandetext.blogspot.com/2011/01/birth-of-photon.html>.
- [4] Wolfe, W. L. and Zissis, G., *The Infrared Handbook*, Office of Naval Research, US Navy, Infrared Information and Analysis Center, Environmental Research Institute of Michigan (1978).
- [5] Pinson, L. J., *Electro-Optics*, John Wiley & Sons, New York (1985) [doi: ISBN 0-471-88142-2].
- [6] Wyatt, C. L., *Radiometric Calibration: Theory and Methods*, Academic Press, New York (1978).
- [7] Boyd, R. W., *Radiometry and the Detection of Optical Radiation*, John Wiley & Sons, New York (1983).
- [8] Bayley, F. J., Owen, J. M., and Turner, A. B., *Heat Transfer*, Nelson Publishers, London (1972).
- [9] Nicodemus, F. E., *NIST Self-Study Manual on Optical Radiation Measurements*, NIST, Washington, D.C. (1976).

- [10] Wyatt, C. L., *Radiometric System Design*, Macmillan Publishing Company, New York (1987).
- [11] "International Lighting Vocabulary," Tech. Rep. CIE No. 17 (E-1.1), Bureau Centrale de la CIE (1970).
- [12] "International Lighting Vocabulary," Tech. Rep. 17.4-1987, Bureau Centrale de la CIE, Bureau Centrale de la CIE (1987).
- [13] "American National Standard Nomenclature and Definitions for illuminating Engineering," Tech. Rep. ANSI Z7.1-1967, Illuminating Engineering Society (1967).
- [14] "American National Standard Nomenclature and Definitions for illuminating Engineering," Tech. Rep. ANSI/IES RP-16-1980, Illuminating Engineering Society (1981).
- [15] Hudson, R. D., *Infrared System Engineering*, Wiley-Interscience, New York (1969).
- [16] RCA Corporation, *RCA Electro-Optics Handbook*, no. 11 in EOH, Burle (1974).
- [17] Edwards, I., "The nomenclature of Radiometry and Photometry," *Laser & Optronics Magazine* 8(8), 37–42 (August 1989).
- [18] Roberts, D. A., "Radiometry & Photometry: Lab Notes on Units," *Photonics Spectra Magazine* 4, 59–63 (April 1987).
- [19] Dresselhaus, M. S., "Solid State Physics (Four Parts)," <http://web.mit.edu/afs/athena/course/6/6.732/www/texts.html>.
- [20] Palmer, J. M. and Grant, B. G., *The Art of Radiometry*, SPIE Press, Bellingham, WA (2009) [doi: 10.1117/3.798237].
- [21] Accetta, J. S. and Shumaker, D. L., Eds., *The Infrared and Electro-Optical Systems Handbook (8 Volumes)*, ERIM and SPIE Press, Bellingham, WA (1993).
- [22] Palik, E. D., Ed., *Handbook of Optical Constants of Solids*, Academic Press, San Diego, CA (1998).
- [23] Modest, M. F., *Radiative Heat Transfer*, Academic Press, San Diego, CA (2003).
- [24] Colour & Vision Research Laboratory, "Colour and Vision Database," <http://www.cvrl.org/index.htm>.

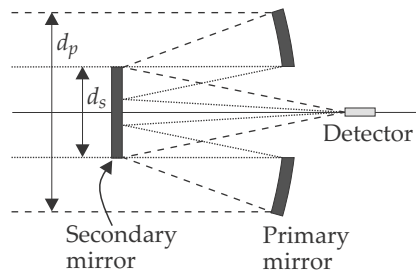
- [25] Pyradi team, "Pyradi data," <https://code.google.com/p/pyradi/source/browse>.
- [26] Schanda, J., Ed., *Colorimetry: Understanding the CIE System*, Wiley-Interscience, New York (2007).
- [27] Wikipedia, "Planckian locus," http://en.wikipedia.org/wiki/Planckian_locus.
- [28] Wikipedia, "CIE 1931 color space," http://en.wikipedia.org/wiki/CIE_1931_color_space.
- [29] Kirk, R., "Standard Colour Spaces," Technical Note FL-TL-TN-0139-StdColourSpaces, FilmLight Digial Film Technology (2007).
- [30] Wikipedia, "Standard illuminant," http://en.wikipedia.org/wiki/Standard_illuminant.
- [31] Wikipedia, "White point," http://en.wikipedia.org/wiki/White_point.

Problems

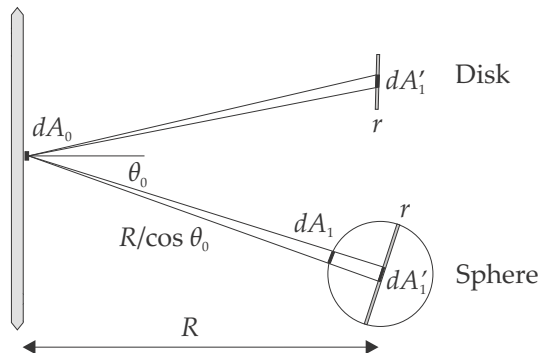
- 2.1 When light interacts with an optical medium on a macroscopic level, there are three potential effects. Describe the three effects that a medium can have on a beam of light and show the mathematical relationship between these effects. [2]
- 2.2 Given the two equations $4\pi \sin^2\left(\frac{\theta}{2}\right)$ and $\pi \sin^2\theta$, explain what they mean and which is which. [2]

Calculate the geometric solid angle and projected solid angle for the following half-apex angles: $\theta \in \{0, 0.01, 0.1, 1, \pi/2\}$ rad. Compare these values in a table and explain why they are different. [2]
- 2.3 A room has floor dimensions of 5 m by 5 m. The roof is 3.5 m above the floor. There is a sensor mounted exactly in the center of the roof. Calculate both the geometric and projected solid angles of the floor as seen from the location of the sensor. Draw a picture of the room, showing all of the details. Derive a mathematical description of the solid angle. Calculate the solid angle numerically (not analytically). [10]
- 2.4 Starting from first principles, derive the solid angle through the optical aperture (opening) of a Cassegrain telescope, as seen from

the detector in the focal plane. Provide all steps of all mathematical derivations. The telescope is as follows:



- 2.5 Describe, in graphical and mathematical terms, how the flux transfer between two arbitrary objects can be calculated. [4]
- 2.6 A circular disk and sphere of radii r are located at a distance R from an infinitely large wall, with uniform radiance L . The normal vector from the disk is normal to the wall. Derive mathematical formulations that describe the flux flowing from the wall through the disk and sphere. [6]



- 2.7 Show mathematically that the irradiance from a Lambertian radiating sphere with radius r is the same as the irradiance of a perpendicularly viewed, Lambertian, radiating circular disk of the same radius (viewed from the same distance) when the distance approaches infinity $R \rightarrow \infty$. To assist in the answer, construct a 'measurement' sphere around the receiving area and consider the projected areas onto this 'measurement' sphere from both the disk and the radiating sphere. [8]

Explain why the sun and moon seem to have a uniform radiance when viewed from earth, even though these bodies are (approximately) spherical. [2]

Perform an Internet search and confirm whether the above state-

ment is true. Support your conclusion with your research findings. [2]

- 2.8 A sphere with a radius r is a distance R from an infinitely large wall, with uniform radiance L . Compute the flux transfer at a given wavelength λ between the sphere and the wall. Do not do the calculation numerically; only show the solution in mathematical terms. [4]

Extend the solution above and show how the flux transferred in the spectral band λ_1 to λ_2 can be calculated if the medium between the sphere and wall has a transmittance of τ_λ . [2]

- 2.9 Considering the results in Figure 2.19, the color value (darkness) of the sample does not seem to affect the color coordinates.

- 2.9.1 Although not shown in the graph, it could well be that a dark-green leaf has the same color coordinates as an apparently lighter-green leaf. Consider the concept that artists call the 'value' of a color. To what extent are the color coordinates affected by the value of the color? [2]

- 2.9.2 Review Equations (2.57) and (2.60) and consider the effect of absolute irradiance on a sample on the color coordinate of that sample. Do the color coordinates change with respect to illumination level? Support your answer with proper mathematical derivation. To what extent will light leakage into the sample affect the measurement of absolute spectral reflectance? What effect will this have on the color coordinates of the sample? [4]

- 2.10 A cylindrical object has a diameter-to-length aspect ratio of 1:10. Derive a mathematical formulation of the object's geometrical solid angle as function of aspect angle, in the plane containing the rotational centerline of the object. You may assume the object distance to be more than 1000 times the cylinder's diameter. Calculate and plot the solid angle at 5-deg intervals around the object. [10]

- 2.11 Derive a mathematical formulation for the solid angle of a sphere with radius r as seen from a distance d measured from the sphere's center. Plot the sphere's solid angle for this distance d ranging from 0 to $10r$. [3]

Chapter 3

Sources

*There could be no fairer destiny for any physical theory
than that it should point the way to a more comprehensive theory
in which it lives on as a limiting case.*

Albert Einstein

3.1 Planck Radiators

The Planck radiation law is derived in detail in several references.¹⁻³ The discussion presented here aims to convey key insights, rather than rigorous mathematical terminology.

Physical matter (atoms and molecules) at nonzero absolute temperature ($T \neq 0$ K) emits and absorbs electromagnetic radiation. Under thermodynamic equilibrium, the incident electromagnetic field and the atoms are in continual energy exchange, mutually sustaining each other's energy state by photon emission and absorption. If either the electromagnetic field intensity or the object's temperature should change, the energy exchange will adjust until thermodynamic equilibrium is re-established.

Perfect thermodynamic equilibrium could exist in an enclosed cavity with the homogenous walls at uniform temperature, such as inside a hollow ball [shown in Figure 3.1(a)]. As shown in the figure, there are two sources of energy: the radiation field [indicated by (1)] and the cavity walls [indicated by (2)]. At any single frequency ν the electromagnetic radiation field is sustained by the wall photon emission [indicated by (3)] at a single energy transition $h\nu$ at atomic level. Likewise, atoms radiate photons at a specific frequency corresponding with the energy transition in the atom, $h\nu$. The wall's kinetic temperature T is sustained by absorption of photons from the electromagnetic field [indicated by (4)].

Thermodynamic equilibrium means that there is zero energy flow \dot{Q} and zero mass flow \dot{m} across the system's boundary [indicated by (5)]. Hence, the only energy exchange is between the radiation field and the

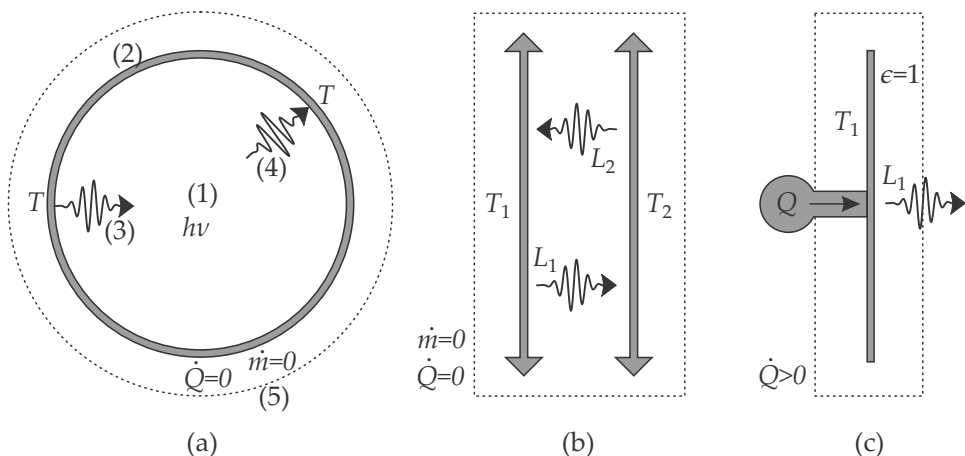


Figure 3.1 Concepts behind Planck's law and thermal radiation: (a) closed cavity, (b) initial non-equilibrium, and (c) permanent non-equilibrium.

cavity walls. Einstein showed⁴ that the principle of detailed balance requires that the processes of spontaneous emission, stimulated emission, and photo-absorption [indicated by (3) and (4)] are in equilibrium. These three processes ensure equilibrium at a single frequency for a given photon but also collectively for all photons at all frequencies.

The permitted energy at each radiative and absorptive frequency must be multiples of the photon's energy $h\nu$, produced by the transition between two energy states in the atoms or molecules. The relative numbers of the allowed energy transitions are, in turn, related by the Boltzmann probability distribution $p(n) = (1 - \exp(-x)) \exp(-nx)$, where $x = h\nu/(kT)$, which is a function of temperature T . The Boltzmann probability distribution provides the link between the wall temperature T and the photon frequency (spectral) distribution.

The final concept in the thermal radiator discussion involves the resonance modes (standing waves) supported in the cavity, also known as the density of states. Density of states is also used in the derivation of the electronic wave function in a crystal (Section 5.5.3) and thermal noise (Section 5.3.2). Density of states is beyond the scope of the current discussion. For more details see complete derivations.¹⁻³

Planck's law applies to transition energy levels compliant with the Boltzmann probability distribution and thus does not apply to lasers, LEDs, fluorescence, or radioactivity (transition levels not compliant with Boltzmann probability distribution). Planck-law radiation is isotropic, spatially homogeneous, unpolarized, and incoherent.

The derivation of Planck's law required an enclosed cavity to retain equilibrium between the electromagnetic wave and the energy in the cavity wall. If a closed system starts from disequilibrium, it will reach equilibrium by exchange of electromagnetic radiation. Consider Figure 3.1(b) with two infinitely large plates at initial temperatures $T_1 > T_2$. The two plates initially emit radiance $L_1 > L_2$. If there is no energy change or mass change in the system, there is a net energy flow only from plate 1 to plate 2 until $T_1 = T_2$ and $L_1 = L_2$, when equilibrium is reached.

Most real-world sources are neither enclosed cavities with uniform temperature nor closed systems with zero mass and energy flow. Figure 3.1(c) shows a plate radiating with radiance L_1 while the plate temperature is maintained at temperature T_1 by an external heat source. The energy supplied by the heat source balances the heat lost by radiation, maintaining constant temperature. This is an example of a laboratory instrument called a blackbody simulator. If the surface emissivity (see Section 3.2) is unity, the surface radiates with a radiance equal to the ideal Planck-law radiator. In this case thermodynamic equilibrium is not required because the heat source maintains the surface temperature at the appropriate temperature. High-performance laboratory blackbody simulators also achieve high emissivity, as described in Section 3.2.5.

Under thermodynamic equilibrium, Planck's law sets the upper limit for an object's radiation for a Boltzmann probability distribution compliant radiator. A Planck-law radiator has an emissivity (see Section 3.2) of unity and is known as a blackbody. Although the blackbody is a theoretical concept, it forms the basis of a surprisingly good model for many real-world objects.

Planck's law can be written in several forms, the spectral variable can be wavelength λ (with units [m] or [μm]), wavenumber $\tilde{\nu}$ (with units [cm^{-1}]), or optical frequency ν (with units [Hz]); and the exitance can be expressed in radiant (watts) or photon (quanta) terms.⁵ Planck's law states that exitance varies with temperature as calculated at a given spectral value. The spectral value is a 'parameter' of Planck's law, and the absolute or kinetic temperature of the object is the only free variable in the law.

In the description of Planck's law, it is convenient to define a variable $x = h\nu/(kT) = hc/(\lambda kT)$ with SI units of [$\text{W}/(\text{m}^3 \cdot \text{K})$], where two of the distance dimensions ($[\text{m}^2]$) relate to area, and one distance dimension ($[\text{m}]$) relates to wavelength. It is also convenient to define a set of radiation constants $c_1 = 2\pi hc^2$ and $c_2 = hc/k$. The constant c_1 has different numerical values depending on the chosen radiation units and the spectral domain variable. These and other physical constants are given in Table A.2.

3.1.1 Planck's radiation law

3.1.1.1 Planck's law in terms of wavelength

The spectral radiant exitance of a blackbody as a function of wavelength is given by⁵

$$M_{e\lambda}(T) = \frac{2\pi hc^2}{\lambda^5 (e^{hc/(\lambda kT)} - 1)} = \frac{c_{1e\lambda}}{\lambda^5 (e^{c_{2\lambda}/(\lambda T)} - 1)}, \quad (3.1)$$

where T is temperature in [K], and λ is wavelength in [m]. Exitance $M_{e\lambda}(T)$ is in units of [W/m²] or [W/(m²·μm)] (depending on the value of $c_{1e\lambda}$). This is a spectral exitance in watts per square meter, per wavelength interval. The values of $c_{1e\lambda}$ and $c_{2\lambda}$ are given in Table A.2.

The derivative with respect to temperature of spectral radiant exitance, for a given temperature T , as a function of wavelength λ , with $x = c_{2\lambda}/(\lambda T)$, is given by

$$\frac{dM_{e\lambda}(T)}{dT} = \frac{2\pi hc^2 x e^x}{\lambda^5 T (e^x - 1)^2} = \frac{c_{1e\lambda} x e^x}{\lambda^5 T (e^x - 1)^2}, \quad (3.2)$$

with units of [W/(m²·K)], or [W/(m²·μm·K)] (depending on the value of $c_{1e\lambda}$). This is a change in spectral exitance in watts per square meter, per wavelength interval, with temperature.

The spectral photon rate exitance at a temperature T as a function of wavelength λ , with $x = c_{2\lambda}/(\lambda T)$, is given by

$$M_{q\lambda}(T) = \frac{2\pi c}{\lambda^4 (e^{hc/(\lambda kT)} - 1)} = \frac{c_{1q\lambda}}{\lambda^4 (e^{c_{2\lambda}/(\lambda T)} - 1)}, \quad (3.3)$$

with exitance in units of [q/(s·m²)] or [q/(s·m²·μm)] (depending on the value of $c_{1q\lambda}$). This is a spectral exitance in photons per second, per square meter, per wavelength interval.

The derivative with respect to temperature of spectral photon rate exitance, for a given temperature T , as a function of wavelength λ , with $x = c_{2\lambda}/(\lambda T)$, is given by

$$\frac{dM_{q\lambda}(T)}{dT} = \frac{2\pi c x e^x}{T \lambda^4 (e^x - 1)^2} = \frac{c_{1q\lambda} x e^x}{T \lambda^4 (e^x - 1)^2}, \quad (3.4)$$

with units of [q/(s·m²·K)] or [q/(s·m²·μm·K)] (depending on the value of $c_{1q\lambda}$). This is a change in spectral exitance in photons per second, per square meter, per wavelength interval, with temperature.

3.1.1.2 Planck's law in terms of wavenumber

The spectral radiant exitance as a function of wavenumber $\tilde{\nu}$ of a blackbody at temperature T is given by

$$M_{e\tilde{\nu}}(T) = \frac{2\pi hc^2 \tilde{\nu}^3}{(e^{ch\tilde{\nu}/(kT)} - 1)} = \frac{c_{1e\tilde{\nu}} \tilde{\nu}^3}{(e^{c_{2\tilde{\nu}} \tilde{\nu}/T} - 1)}, \quad (3.5)$$

with units $[W/(m^2 \cdot m^{-1})]$ or $[W/(m^2 \cdot cm^{-1})]$ (depending on the value of $c_{1e\tilde{\nu}}$). This is a spectral exitance in watts per square meter, per wavenumber interval.

The derivative with respect to temperature of spectral radiant exitance, for a given temperature T , as a function of wavenumber $\tilde{\nu}$, with $x = c_{2\tilde{\nu}} \tilde{\nu}/T$, is given by

$$\frac{dM_{e\tilde{\nu}}(T)}{dT} = \frac{2\pi hc^2 \tilde{\nu}^3 x e^x}{T (e^x - 1)^2} = \frac{c_{1e\tilde{\nu}} \tilde{\nu}^3 x e^x}{T (e^x - 1)^2}, \quad (3.6)$$

with units $[W/(m^2 \cdot m^{-1} \cdot K)]$ or $[W/(m^2 \cdot cm^{-1} \cdot K)]$ (depending on the value of $c_{1e\tilde{\nu}}$). This is a change in spectral exitance in watts, per square meter, per wavenumber interval, with temperature.

The spectral photon exitance as a function of wavenumber ν of a blackbody at temperature T is given by

$$M_{q\tilde{\nu}}(T) = \frac{2\pi c \tilde{\nu}^2}{(e^{ch\tilde{\nu}/(kT)} - 1)} = \frac{c_{1q\tilde{\nu}} \tilde{\nu}^2}{(e^{c_{2\tilde{\nu}} \tilde{\nu}/T} - 1)}, \quad (3.7)$$

with units $[q/(s \cdot m^2 \cdot m^{-1})]$ or $[q/(s \cdot m^2 \cdot cm^{-1})]$ (depending on the value of $c_{1q\tilde{\nu}}$). This is a spectral exitance in photons per second, per square meter, per wavenumber interval.

The derivative with respect to temperature of spectral photon exitance, for a given temperature T , as a function of wavenumber $\tilde{\nu}$, with $x = c_{2\tilde{\nu}} \tilde{\nu}/T$ is given by

$$\frac{dM_{q\tilde{\nu}}(T)}{dT} = \frac{2\pi c \tilde{\nu}^2 x e^x}{T (e^x - 1)^2} = \frac{c_{1q\tilde{\nu}} \tilde{\nu}^2 x e^x}{T (e^x - 1)^2}, \quad (3.8)$$

with units $[q/(s \cdot m^2 \cdot m^{-1} \cdot K)]$ or $[q/(s \cdot m^2 \cdot cm^{-1} \cdot K)]$ (depending on the value of $c_{1q\tilde{\nu}}$). This is a change in spectral exitance in photons per second, per square meter, per wavenumber interval, with temperature.

3.1.1.3 Planck's law in terms of frequency

The spectral radiant exitance as a function of frequency ν of a blackbody at temperature T is given by

$$M_{ev}(T) = \frac{2\pi h \nu^3}{c^2 \left(e^{\frac{h\nu}{kT}} - 1 \right)} = \frac{c_{1ev} \nu^3}{(e^{c_{2\nu} \nu/T} - 1)}, \quad (3.9)$$

with units $[W/(m^2 \cdot Hz)]$. This is a spectral exitance in watts per square meter, per frequency interval.

The derivative with respect to temperature of radiant exitance, for a given temperature T , as a function of frequency ν , with $x = c_{2\nu} \nu/T$, is given by

$$\frac{dM_{ev}(T)}{dT} = \frac{2\pi h \nu^3 x e^x}{c^2 T \left(e^{\frac{h\nu}{kT}} - 1 \right)^2} = \frac{c_{1ev} \nu^3 x e^x}{T (e^x - 1)^2}, \quad (3.10)$$

with units $[W/(m^2 \cdot Hz \cdot K)]$. This is a change in spectral exitance in watts per square meter, per frequency interval, with temperature.

The spectral photon rate exitance as a function of frequency ν of a blackbody at temperature T is given by

$$M_{qv}(T) = \frac{2\pi \nu^2}{c^2 \left(e^{\frac{h\nu}{kT}} - 1 \right)} = \frac{c_{1qv} \nu^2}{(e^{c_{2\nu} \nu/T} - 1)}, \quad (3.11)$$

with units $[q/(s \cdot m^2 \cdot Hz)]$. This is a spectral exitance in photons per second, per square meter, per frequency interval.

The derivative with respect to temperature of spectral photon exitance, for a given temperature T , as a function of frequency ν , with $x = c_{2\nu} \nu/T$, is given by

$$\frac{dM_{qv}(T)}{dT} = \frac{2\pi \nu^2 x e^x}{c^2 T (e^x - 1)^2} = \frac{c_{1qv} \nu^2 x e^x}{T (e^x - 1)^2}, \quad (3.12)$$

with units $[q/(s \cdot m^2 \cdot Hz \cdot K)]$. This is a change in spectral exitance in photons per second, per square meter, per frequency interval, with temperature.

3.1.2 Wien's displacement law

From Figure 3.2 it is clear that the Planck-law radiation curve has only one maximum. The equation relating the blackbody temperature and the spectral value at the peak exitance is known as Wien's displacement law. The

spectral value (wavelength, frequency, wavenumber) where the maximum exitance occurs is obtained by differentiating Planck's law with respect to the spectral variable, equating it to zero, and solving for the spectral variable.⁶ Note that the different maxima do not coincide.

The relationship between the blackbody temperature in [K] and the spectral variable (λ_{me} in [μm], $\tilde{\nu}_{me}$ in [cm^{-1}], or ν_{me} in [Hz]) at which maximum radiant exitance occurs is given by

$$\lambda_{me} = \frac{10^6 hc}{a_5 k T} = \frac{w_{e\lambda}}{T}, \quad (3.13)$$

$$\tilde{\nu}_{me} = \frac{a_3 k T}{100 hc} = w_{e\tilde{\nu}} T, \quad (3.14)$$

and

$$\nu_{me} = \frac{a_3 k T}{h} = w_{e\nu} T. \quad (3.15)$$

The relationship between the blackbody temperature in [K] and the spectral variable (λ_{mq} in [μm], $\tilde{\nu}_{mq}$ in [cm^{-1}], or ν_{mq} in [Hz]) at which the maximum photon rate exitance occurs and is given by

$$\lambda_{mq} = \frac{10^6 hc}{a_4 k T} = \frac{w_{q\lambda}}{T}, \quad (3.16)$$

$$\tilde{\nu}_{mq} = \frac{a_2 k T}{100 hc} = w_{q\tilde{\nu}} T, \quad (3.17)$$

and

$$\nu_{mq} = \frac{a_2 k T}{h} = w_{q\nu} T. \quad (3.18)$$

3.1.3 Stefan–Boltzmann law

If Planck's law is integrated over all wavelengths, the total radiant exitance from a blackbody is obtained:³

$$M_e(T) = \frac{2k^4\pi^5}{15c^2h^3} T^4 = \sigma_e T^4, \quad (3.19)$$

with exitance $M_e(T)$ in [W/m^2], σ_e the Stefan–Boltzmann constant in units of [$\text{W}/(\text{m}^2 \cdot \text{K}^4)$], and temperature T in [K]. Note that the Stefan–Boltzmann law does not consider energy balance between incident flux and radiated flux — it assumes the environment is at 0 K, with no incident flux.

The total photon rate exitance from a blackbody is⁶

$$M_q(T) = \frac{4\zeta(3)k^3}{h^3c^2}T^3 = \sigma_q T^3, \quad (3.20)$$

with exitance $M_q(T)$ in $[\text{q}/(\text{s}\cdot\text{m}^2)]$, σ_q the Stefan–Boltzmann constant in $[\text{q}/(\text{s}\cdot\text{m}^2\cdot\text{K}^3)]$, and temperature T in $[\text{K}]$.

3.1.4 Summation approximation of Planck's law

Planck's law can be written in the form of an infinite sum.⁵ Consider the integral of the Planck radiation law as a function of wavelength. Starting from first principles, where $c_1 = 2\pi hc^2$, $c_2 = hc/k$, and $x = c_2/(\lambda T)$,

$$\begin{aligned} L &= \int_{\lambda_1}^{\lambda_2} \frac{2c^2h}{\lambda^5} \frac{d\lambda}{(e^x - 1)} \\ &= \int_{\lambda_1}^{\lambda_2} \frac{c_1}{\pi\lambda^5} \frac{d\lambda}{(e^x - 1)}. \end{aligned} \quad (3.21)$$

Change the integration variable $\lambda \rightarrow c_2/(xT)$ and $d\lambda \rightarrow -(c_2 dx)/(Tx^2)$:

$$\begin{aligned} L &= \frac{c_1}{\pi} \int_{x_2}^{x_1} \left(\frac{xT}{c_2} \right)^5 \frac{c_2}{Tx^2} \frac{dx}{e^x - 1} \\ &= \frac{c_1 T^4}{\pi c_2^4} \int_{x_2}^{x_1} \frac{x^3}{e^x - 1} dx \\ &= \frac{c_1 T^4}{\pi c_2^4} \int_{x_2}^{x_1} x^3 \sum_{m=1}^{\infty} e^{-mx} dx \\ &= \frac{c_1 T^4}{\pi c_2^4} \sum_{m=1}^{\infty} \int_{x_2}^{x_1} x^3 e^{-mx} dx, \end{aligned} \quad (3.22)$$

where $e^x - 1$ is expanded into the infinite sum through long division. Applying integral number 2.322.3 in Gradshteyn,⁷

$$\begin{aligned} L &= \frac{c_1 T^4}{\pi c_2^4} \sum_{m=1}^{\infty} e^{-mx} \left(-\frac{x^3}{m} - \frac{3x^2}{m^2} - \frac{6x}{m^3} - \frac{6}{m^4} \right) \Bigg|_{x_2}^{x_1} \\ &= \frac{c_1 T^4}{\pi c_2^4} \sum_{m=1}^{\infty} \frac{e^{-mx}}{m^4} ((xm)^3 + 3(xm)^2 + 6xm + 6) \Bigg|_{x_1}^{x_2}. \end{aligned} \quad (3.23)$$

No approximations are made in the derivation, and the formula is therefore exact, provided that enough terms are used in the summation. Because each value of m results in ten free parameters, it implies that the number of free parameters increases rapidly for increasing m , limiting the useful application of Equation (3.23) unless the algorithm is coded on a computer.

3.1.5 Summary of Planck's law

The Planck-law function and Planck-law temperature derivative are summarized in Table 3.1 and plotted for several temperatures in Figures 3.2 and 3.3. Note in Figure 3.2 the two important properties of blackbody radiation: with an increase in temperature, the exitance increases rapidly and the exitance peak shifts toward shorter wavelengths. Also note the very large rate of photons emitted, even by an object at 0 °C (273.15 K).

The constants in Tables 3.1 and A.2 were calculated from the International Council for Science: Committee on Data for Science and Technology (CODATA)⁸ constants as encoded in SciPy.⁹ Python code implementations of the Planck-law equations are available in Section D.4.1 and in the pyradi¹⁰ toolset.

3.1.6 Thermal radiation from common objects

Figure 3.4 provides a practical view of blackbody radiation — assuming unity emissivity. The top graph shows the radiance for several common objects. The bottom graph shows the normalized cumulative radiance for the same objects, where the cumulative spectral radiance is normalized by the Stefan–Boltzmann equation. Note that the 50% cumulative radiance value occurs at a surprisingly longer wavelength than the peak radiance. For example, it is evident that a long-wave infrared camera captures approximately 30% of the total radiance of a 300-K object even though the peak radiance occurs at 10 μm . Also shown in the figure are commonly used spectral band designations. These band designations come from transparent ‘windows’ in the atmosphere (see Section 4.6.4) and the traditional availability of optical detector materials, providing sensitivity in the various bands (see Chapter 5). These designations are not well defined or standardized, and only serve as a rough indication of spectral band. The acronyms are as follows: near-infrared (NIR, 0.75–1.4 μm), short-wave infrared (SWIR, 1.5–2.5 μm), medium-wave infrared (MWIR, 3–5 μm), and long-wave infrared (LWIR, 8–12 μm).

3.2 Emissivity

This section covers the emissivity concept. The work started here is expanded in Section 6.6, which considers the modeling of thermal radiators. The Planck radiator (or blackbody) is a very convenient basis for modeling a large class of sources. However, none of these sources behaves exactly like a blackbody radiator. In order to use Planck's law with real sources, it

Table 3.1 Planck's law summary (constants also in Table A.2).

Wavelength	Wavenumber
λ in [μm], T in [K] $x = c_{2\lambda}/(\lambda T) = 14387.8/(\lambda T)$ $c_{1e\lambda} = 3.74177 \times 10^8$ $c_{1q\lambda} = 1.88365 \times 10^{27}$	$\tilde{\nu}$ in [cm^{-1}], T in [K] $x = c_{2\tilde{\nu}} \tilde{\nu}/T = 1.43878 \tilde{\nu}/T$ $c_{1e\tilde{\nu}} = 3.74177 \times 10^{-8}$ $c_{1q\tilde{\nu}} = 1.88365 \times 10^{15}$
Planck's law	
$M_{e\lambda} = \frac{c_{1e\lambda}}{\lambda^5 (e^x - 1)}$ in [$\text{W}/(\text{m}^2 \cdot \mu\text{m})$]	$M_{e\tilde{\nu}} = \frac{c_{1e\tilde{\nu}} \tilde{\nu}^3}{(e^x - 1)}$ in [$\text{W}/(\text{m}^2 \cdot \text{cm}^{-1})$]
$M_{q\lambda} = \frac{c_{1q\lambda}}{\lambda^4 (e^x - 1)}$ in [$\text{q}/(\text{s} \cdot \text{m}^2 \cdot \mu\text{m})$]	$M_{q\tilde{\nu}} = \frac{c_{1q\tilde{\nu}} \tilde{\nu}^2}{(e^x - 1)}$ in [$\text{q}/(\text{s} \cdot \text{m}^2 \cdot \text{cm}^{-1})$]
Temperature derivative of Planck's law	
$\frac{dM_{e\lambda}}{dT} = \frac{c_{1e\lambda} x e^x}{T \lambda^5 (e^x - 1)^2}$ in [$\text{W}/(\text{m}^2 \cdot \mu\text{m} \cdot \text{K})$]	$\frac{dM_{e\tilde{\nu}}}{dT} = \frac{c_{1e\tilde{\nu}} \tilde{\nu}^3 x e^x}{T (e^x - 1)^2}$ in [$\text{W}/(\text{m}^2 \cdot \text{cm}^{-1} \cdot \text{K})$]
$\frac{dM_{q\lambda}}{dT} = \frac{c_{1q\lambda} x e^x}{T \lambda^4 (e^x - 1)^2}$ in [$\text{q}/(\text{s} \cdot \text{m}^2 \cdot \mu\text{m} \cdot \text{K})$]	$\frac{dM_{q\tilde{\nu}}}{dT} = \frac{c_{1q\tilde{\nu}} \tilde{\nu}^2 x e^x}{T (e^x - 1)^2}$ in [$\text{q}/(\text{s} \cdot \text{m}^2 \cdot \text{cm}^{-1} \cdot \text{K})$]
Wien's displacement law	
$\lambda_{me} = 2897.77212/T$	$\tilde{\nu}_{me} = T \times 1.960998438$
$\lambda_{mq} = 3669.7031/T$	$\tilde{\nu}_{mq} = T \times 1.107624256$
Stefan–Boltzmann law	
$M_e(T) = 5.670373 \times 10^{-8} T^4$ in [W/m^2]	$M_q(T) = 1.5204606 \times 10^{15} T^3$ in [$\text{q}/(\text{s} \cdot \text{m}^2)$]

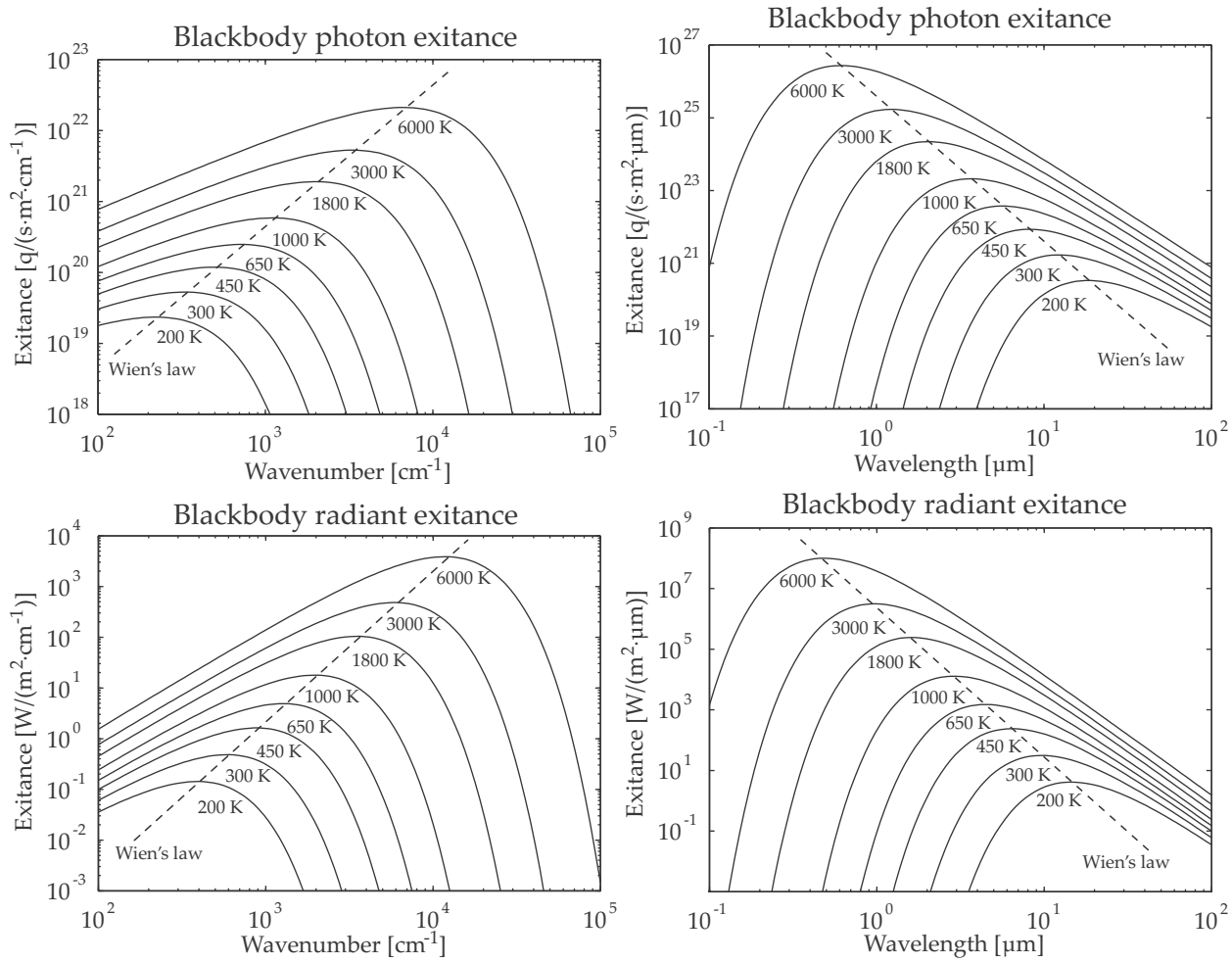


Figure 3.2 Planck's law and Wien's displacement law for various temperatures.

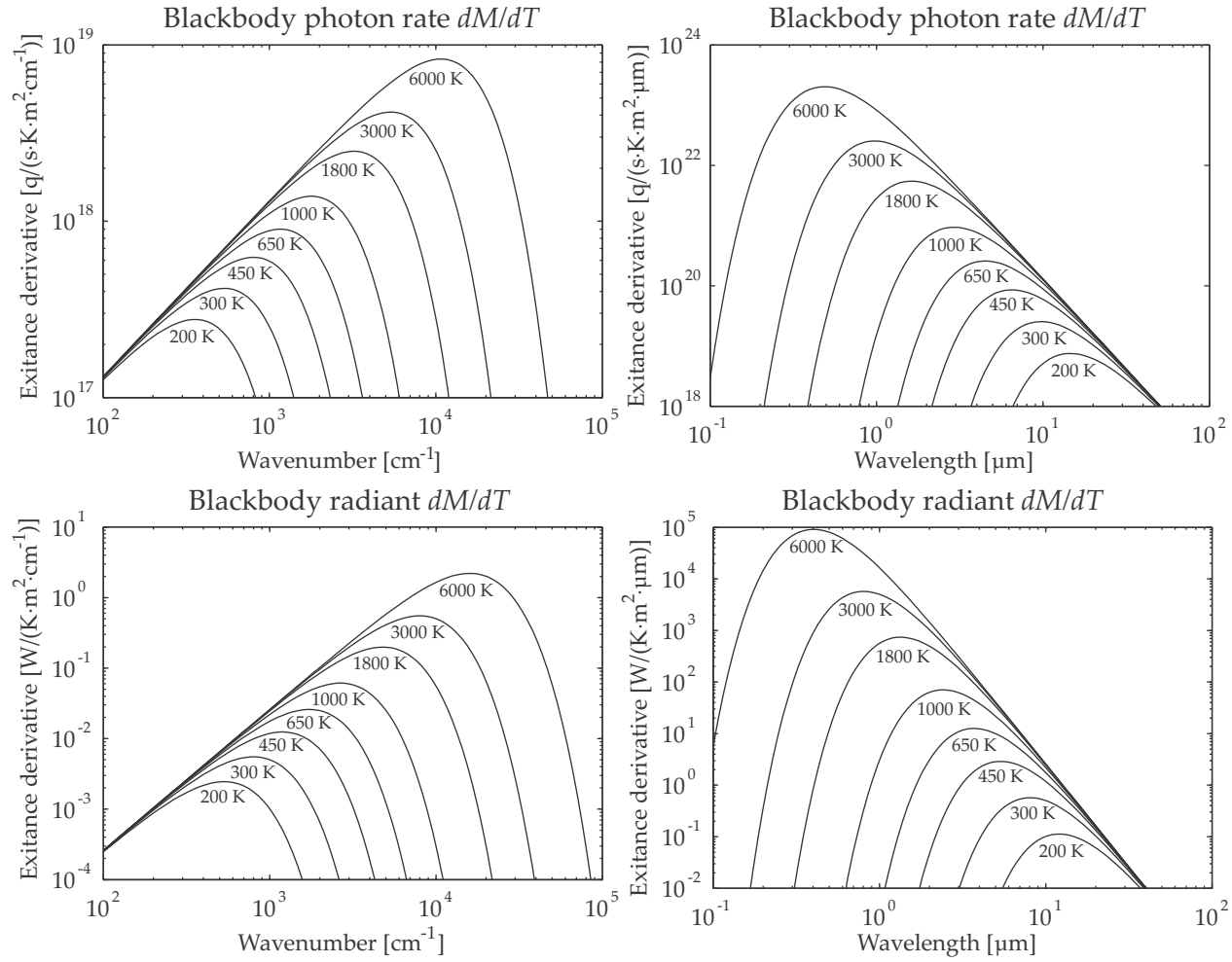


Figure 3.3 Temperature derivative of Planck's law for various temperatures.

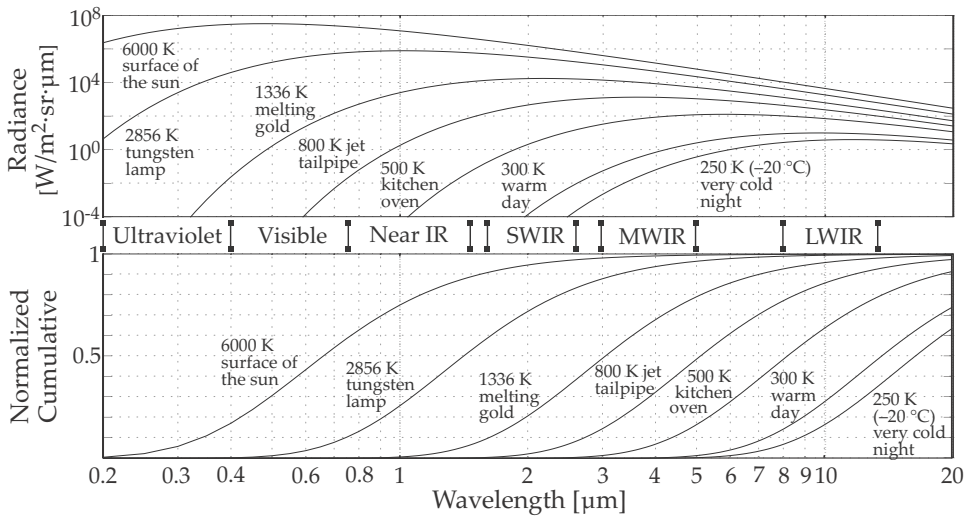


Figure 3.4 Summary blackbody curves of common objects.

is convenient to define emissivity ϵ as *the degree to which a thermal radiator approximates a blackbody*:

$$\epsilon_{\lambda} = \frac{L_{\text{object}\lambda}}{L_{\text{bb}\lambda}}. \quad (3.24)$$

Emissivity has many guises, to allow for spectral and directional parameters, that are necessary to describe emissivity in a more general sense. Table 3.2 provides a short summary of the various definitions of emissivity. The source radiance is indicated by the subscript s , and the theoretical blackbody radiance is indicated by the subscript $_{\text{bb}}$. The source and blackbody temperatures are equal, $T_s = T_{\text{bb}}$. The two variables θ and φ denote directional zenith and azimuth angles, respectively. A much more detailed description is given in Palmer and Grant.¹¹ Directional emissivity and reflectance are investigated in more detail in Section 3.4.

3.2.1 Kirchhoff's law

Kirchhoff's law can be summarized as:^{2,11–13} “For an object in thermodynamic equilibrium with its surroundings, the absorptivity α of an object is exactly equal to its emissivity ϵ , in each direction and at each wavelength $\epsilon_{\lambda}(T, \theta, \varphi) = \alpha_{\lambda}(T, \theta, \varphi)$.” This statement essentially means that a good radiator is also a good absorber.

The statement of equality between emissivity and absorptivity is not unconditional — spectral and angular variations in emissivity may lead to

Table 3.2 Definitions of various forms of emissivity.

Quantity	Symbol	Definition
Spectral directional emissivity	$\epsilon_\lambda(\theta, \varphi)$	$\frac{L_{\lambda s}(\theta, \varphi)}{L_{\text{bb}\lambda}(\theta, \varphi)}$
Directional total emissivity	$\epsilon(\theta, \varphi)$	$\frac{\int_0^\infty L_{\lambda s}(\theta, \varphi) d\lambda}{\int_0^\infty L_{\text{bb}\lambda}(\theta, \varphi) d\lambda}$
Spectral hemispherical emissivity	ϵ_λ	$\frac{\int_0^{\pi/2} \int_0^{2\pi} L_{\lambda s}(\theta, \varphi) \cos \theta d\varphi d\theta}{\int_0^{\pi/2} \int_0^{2\pi} L_{\text{bb}\lambda}(\theta, \varphi) \cos \theta d\varphi d\theta}$
Hemispherical total emissivity	ϵ	$\frac{\int_0^\infty \int_0^{\pi/2} \int_0^{2\pi} L_{\lambda s}(\theta, \varphi) \cos \theta d\varphi d\theta d\lambda}{\int_0^\infty \int_0^{\pi/2} \int_0^{2\pi} L_{\text{bb}\lambda}(\theta, \varphi) \cos \theta d\varphi d\theta d\lambda}$

apparent violations of Kirchhoff's law if not accounted for, as shown in Table 3.2.

In Section 2.3.4 it is stated that $1 = \alpha + \tau + \rho$. For an opaque surface $\tau = 0$, and it follows that $\alpha = \epsilon = 1 - \rho$, which implies that a good absorber/radiator has a low reflectivity. Because a blackbody has an emissivity of unity, the reflectance is zero, resulting in a surface with a black appearance (hence the name blackbody).

For a gaseous radiator $\rho = 0$, and it follows that $\alpha = \epsilon = 1 - \tau$, which implies that a gas medium with low transmittance will radiate with high emissivity. In the extreme case where the transmittance is very low, e.g., 1%, the emissivity is very high, approaching a blackbody radiator.

3.2.2 Flux transfer between a source and receiver

Equation (2.31) describes the flux flowing in a radiance field, passing through both surfaces. As stated, this equation does not consider the origin of the flux nor the destination of the flux. Armed with the concepts of emissivity and absorption, now add a radiator element with emissivity ϵ_0 behind dA_0 and an absorber element with absorption α_1 behind dA_1 , as shown in Figure 3.5. In this construction, the flux leaving from a radiator at the source, being absorbed by another absorber at the receiver, is given by

$$d^2\Phi = \frac{L_{\text{bb}} \epsilon_0 dA_0 \cos \theta_0 \alpha_1 dA_1 \cos \theta_1}{R^2}. \quad (3.25)$$

Equation (3.25) defines only the flux from a small area dA_0 being absorbed in dA_1 . Note that there could be flux from many other small sources also absorbed in the receiver dA_1 .

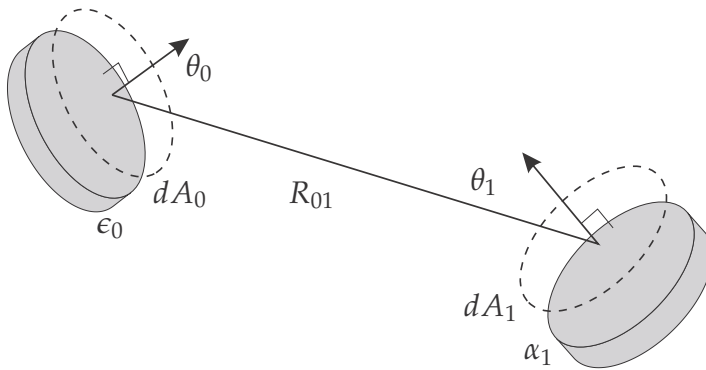


Figure 3.5 Radiative flux between a source and receiver.

3.2.3 Grey bodies and selective radiators

A grey body radiator is a thermal radiator with a spectrally invariant emissivity less than unity. The grey body radiation at any wavelength is therefore a constant fraction of the blackbody radiation at the same temperature. In practice, no physical object is a true grey body as defined above. But if the emissivity is reasonably constant over the spectral range of interest, the object is referred to as a grey body radiator over that spectral band. Examples of grey bodies include most natural objects, with emissivity mostly ranging from 0.5 to 0.99 in the 8–12- μm spectral range.

A selective radiator is a radiator with spectrally variant emissivity. The spectral emissivity varies slowly or very abruptly in the spectral range of interest. Examples of selective radiators include molecular gas emission lines, the wavelengths of which are related to the differences in the energy states in the gas and are thus characteristic of the gas composition. These lines can be very narrow and are usually found in clusters.

Figure 3.6 illustrates the concepts of black, grey, and selective radiators. For any thermal radiator, Planck's law sets an upper limit to the radiation emanating from the radiator. Examples of spectral emissivity are shown in Figure 3.7. The water vapor emissivity values were calculated for a 100-m path in a MODTRAN^{TM14} Tropical atmosphere at sea level. The significance of this curve is that even for short paths the water vapor in the atmosphere radiates at near-blackbody radiance at some wavelengths. The CO₂ emissivity curve was measured by placing a Bunsen burner flame near the calibration port of a Fourier transform spectrometer.

Spectral radiator emissivity is an aggregate of many narrow 'lines,'

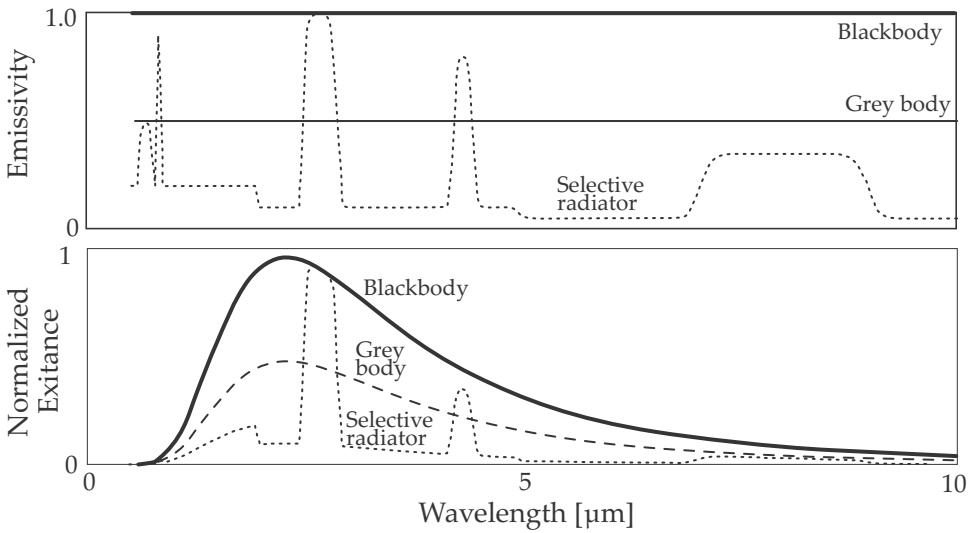


Figure 3.6 Blackbody, grey body, and selective thermal radiators.

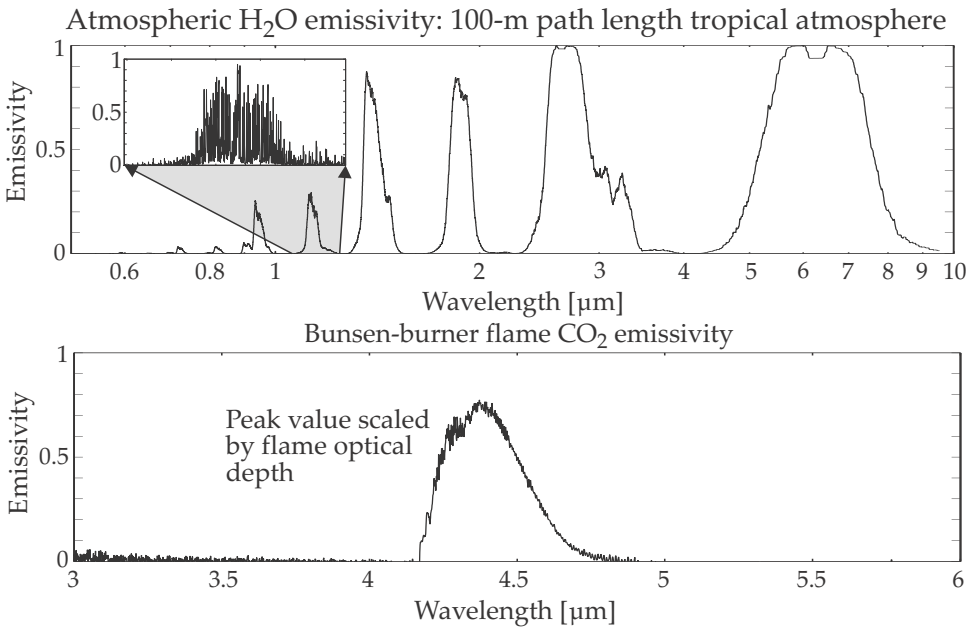


Figure 3.7 Spectral emissivity for H₂O and CO₂.

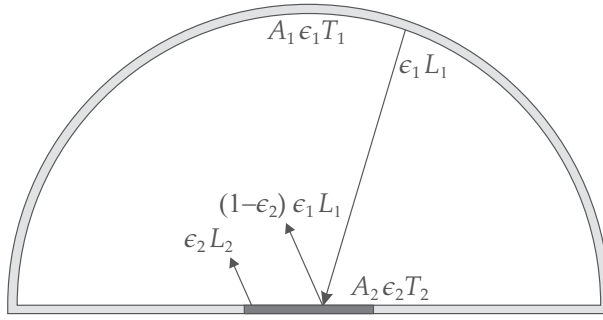


Figure 3.8 Dome enclosing a small component.

where each line corresponds to a discrete energy level in the molecule or atom. All of the molecules are at slightly different temperatures, with the result that the individual lines are also slightly displaced. Figure 3.7 shows an example of the line structure in a small spectral range.

3.2.4 Radiation from low-emissivity surfaces

The self-radiation from a single surface of given temperature depends on the emissivity of the surface. If the emissivity is low, the self-radiation will be concomitantly low. If the surface is opaque, a low emissivity implies a high reflectivity, which in turn means high reflection of the ambient light. Consider the geometry defined in Figure 3.8, with surface 1 having high emissivity ϵ_1 and temperature T_1 , and surface 2 having a low emissivity ϵ_2 and physical temperature T_2 . Assuming Lambertian surfaces, the radiance of surface 2 is

$$L_2 = \epsilon_2 L_e(T_2) + (1 - \epsilon_2) \epsilon_1 L_e(T_1). \quad (3.26)$$

Noncontact infrared temperature measurement estimates an object's temperature, called the 'apparent' or 'radiation' temperature, by comparing the radiance of the object with the radiance of a reference source (normally a blackbody). The user adjusts the emissivity setting in the instrument to match the object's emissivity. This setting can be in error (e.g., if the object emissivity is not known), leading to an incorrect temperature estimate. If the object's emissivity is estimated to be ϵ_{m2} , the apparent or radiation temperature T_{m2} can be determined by solving

$$\epsilon_{m2} \sigma_e T_{m2}^4 = \epsilon_2 \sigma_e T_2^4 + (1 - \epsilon_2) \epsilon_1 \sigma_e T_1^4. \quad (3.27)$$

Even if $\epsilon_{m2} = \epsilon_2$, the apparent temperature measurement is not the physical surface temperature. Most noncontact temperature measurement probes are set to assume a high source emissivity. If it is assumed that $\epsilon_{m2} = 1$

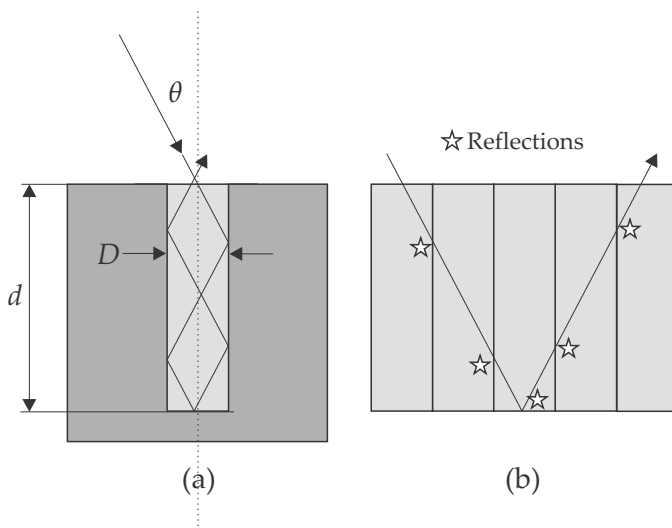


Figure 3.9 Multiple reflections in a cavity: (a) simple air-filled cavity, and (b) reflectance of a light ray entering the graphically unfolded cavity.

and $\epsilon_1 = 1$, the apparent temperature of surface 2 varies between T_1 (when $\epsilon_2 = 0$) and T_2 (when $\epsilon_2 = 1$), as is evident from

$$T_{a2}^4 = \epsilon_2 T_2^4 + (1 - \epsilon_2) T_1^4.$$

In practical terms, this means that a noncontact radiation temperature measurement of a low-emissivity surface requires an accurate estimate of the surface emissivity.

3.2.5 Emissivity of cavities

The definition of an optical cavity is a volume of given refractive index inside a larger volume of different refractive index. Consider a simple air-filled cavity in a solid block as shown in Figure 3.9(a). The reflectance of a light ray entering the cavity is graphically unfolded in Figure 3.9(b). The total reflectance of the cavity is given by ρ^N , where ρ is the reflection at a single surface, and N is the number of reflectance events. Because the block is opaque ($\tau = 0$), the emissivity of the cavity after N reflections is given by $\epsilon = 1 - \rho^N$. If $\rho < 1$, the emissivity increases with increase in the number of reflections. In the limit, when $n \rightarrow \infty$, $\epsilon \rightarrow 1$. Cavities therefore exhibit higher emissivity than that of the surfaces forming the cavity. Cavities can be constructed, such as using a cavity to increase the apparent emissivity of a thermal source; careful cavity design can result in near-unity emissivity. Cavities also occur commonly in nature in the form of micro-cavities on the surface of rough objects (discussed in Section 3.4).

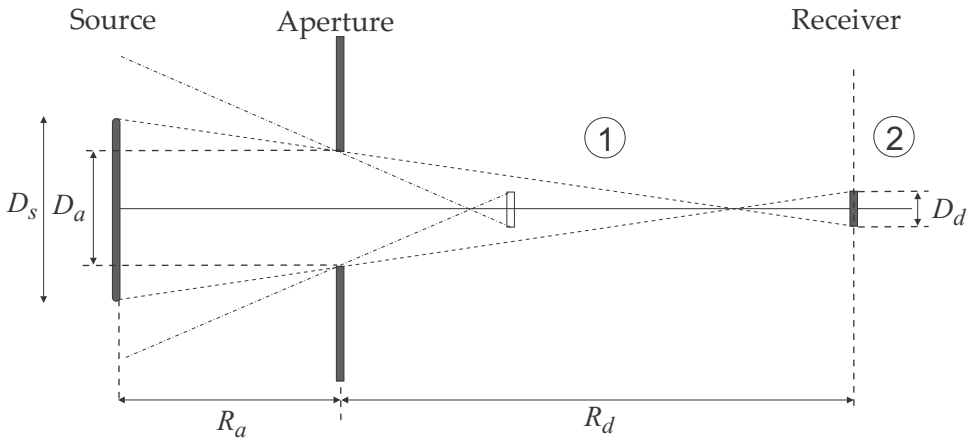


Figure 3.10 Aperture plate beam vignetting.

3.3 Aperture Plate in front of a Blackbody

Laboratory blackbodies are commonly used with well-defined aperture plate diameters to obtain sources with well-defined areas, such as the configuration shown in Figure 3.10. The requirement is that, through the aperture plate, every elemental area of the source must irradiate every elemental area of the receiver. Put differently, when the receiver observes the source through the aperture plate, it must only see some part of the source and nothing beyond the source. If this condition is met, the area of the source becomes irrelevant, and the area of the aperture plate applies. It is as if the source surface is located in the plane of the aperture plate, irrespective of where the physical source is located. This condition occurs as a result of the spatial conservation of radiance in a nonlossy medium: the (assumed uniform) radiance on the source radiating surface is the same as the radiance in the plane of the aperture plate.

3.4 Directional Surface Reflectance

Surface micro-roughness, at the scale of the wavelength of light, has a significant effect on reflection and radiant exitance from the surface. The surface can be considered to consist of a great many micro-facets. Figure 3.11 shows four surface-roughness cases¹⁵ in terms of the ratio of root-mean-square (rms) roughness σ to the wavelength λ of the incident light. The roughness is not only a measure of vertical variation, in most surfaces the horizontal scale also varies with vertical scale; i.e., it is unlikely to have large vertical variations with very small horizontal intervals.

The surface can be smooth [Figure 3.11(a)], or it can be more complex [Figure 3.11(d)]. Such complex surfaces may have semi-transparent multi-layers, each with its own surface irregularities and volumetric properties (e.g., oxides, synthetic or natural thin films), whereas other surfaces could contain stacked particles forming porous volumes (e.g., sandpaper, soot, or dust). Constructed light traps¹⁶ have surfaces with geometric shapes to capture light by means of multiple reflections, aligned such that the light may never escape from the surface.

An optically flat surface, where $\sigma/\lambda \approx 0$ [shown in Figure 3.11(a)], reflects light according to geometrical optics because the near-zero surface roughness has little scattering effect. At the other extreme where $\sigma/\lambda > 1$ [shown in Figure 3.11(d)], the roughness scale exceeds the wavelength of light and the surface micro-facets act as independent small mirrors, each reflecting according to geometrical optics. Between these two extremes [Figures 3.11(b) and (c)], the surface roughness scatters the light in spatial directions ranging from specular (broadened around the mirror vector) to perfectly diffuse (Lambertian). In this region the bidirectional reflection distribution function (BRDF)¹⁷ is used to describe the spatial reflectance. Various BRDF models are used, some based on empirical approximation, whereas others attempt to model rigorously from first principles. In practice the BRDF is most easily obtained by measurement¹⁸ or a combination of measurement and theoretical modeling.^{15,19}

A Lambertian reflector has an important surface property that the reflected (or emitted) *radiance* (what humans perceive as visual brightness) is the same in all directions, irrespective of the direction or source of the incident light. One example of a Lambertian surface is an opaque object with a micro-scale rough and porous surface, where scattering from subsurface roughness dominates.

3.4.1 Roughness and scale

Figure 3.11 assumes an illumination beam size comparable with the microscopic scale of a homogenous surface. A similar principle also applies to aggregate properties of composite surfaces. Consider a satellite camera viewing a crop field. In this case the sun is the illuminating ‘beam,’ and a single satellite-sensor pixel FOV observes an area comprising crop and soil. The camera pixel FOV footprint covers several rows of the crop. The observed radiance will vary depending on the sun-field-camera viewing geometry, i.e., illumination and/or viewing along or across the rows of plants. Thus, depending on the application, roughness is not only expressed in terms of wavelength scale but also in FOV footprint scale.

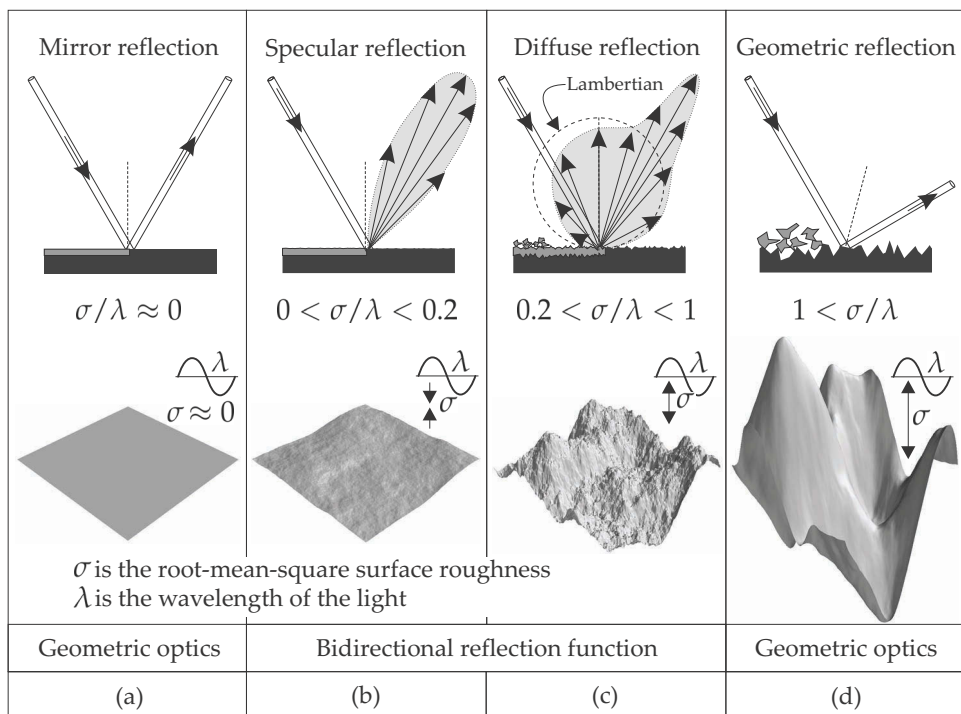


Figure 3.11 Micro-scale surface roughness and reflection.

3.4.2 Reflection geometry

The geometry describing reflection is shown in Figure 3.12. Let the incident ray be defined by the unit vector \hat{I} , the surface normal vector by the unit vector \hat{N} , and the reflected ray vector by the unit vector \hat{R} . It is shown in Appendix C that the direction of the reflected ray is given by $\hat{R} = \hat{I} - 2(\hat{I} \cdot \hat{N})\hat{N}$. In this definition \hat{R} represents the mirror reflection vector from the surface. Consider now the direction of the measured flux as reflected along any arbitrary vector \hat{S} . The vector \hat{S} has an angle α with respect to the mirror reflection vector \hat{R} (where $\cos \alpha = \hat{S} \cdot \hat{R}$) and an angle θ_s with respect to the surface normal vector \hat{N} (where $\cos \theta_s = -\hat{S} \cdot \hat{N}$). The incident ray vector \hat{I} has an angle θ_i with respect to the surface normal vector \hat{N} (where $\cos \theta_i = -\hat{I} \cdot \hat{N}$). The mirror reflected ray vector \hat{R} has an angle θ_r with respect to the surface normal vector \hat{N} (where $\cos \theta_r = \hat{R} \cdot \hat{N}$).

3.4.3 Reflection from optically smooth surfaces

The reflection from an optically smooth surface $\sigma/\lambda \approx 0$ is determined by the material's index of refraction.²⁰ The Fresnel equation describes the

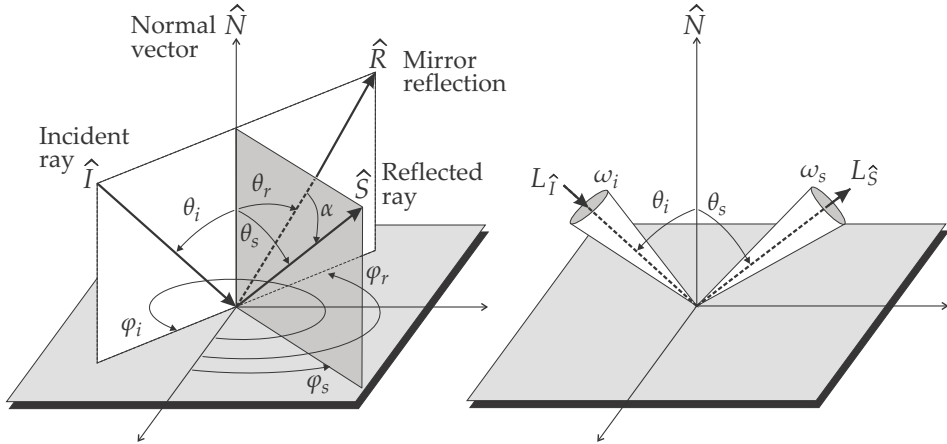


Figure 3.12 Reflection geometry.

reflection as a function of incident angle for both conducting metals and nonconducting dielectric materials. Whereas dielectrics (e.g., water, glass) have real and small indices of refraction (low reflectance), metals have complex and large indices of refraction (high reflectance). The direction of the reflected light from the smooth surface will be the mirror reflection \hat{R} in Figure 3.12.

A conducting medium can be modeled as a gas of unbound charges circulating in the medium (electrons in a metal).²⁰ These free electrons and their accompanying positive nuclei can undergo ‘plasma oscillations’ at a resonant plasma frequency ν_p . The refractive index for a metal is then given by $n_v^2 = 1 - (\nu_p/\nu)^2$ (see Section 5.5.8). When $\nu > \nu_p$, n is real, and the metal becomes transparent. When $\nu < \nu_p$, n is complex, and the imaginary component leads to absorption into the material. However, most of the light energy is not dissipated, and the wave is reflected from the surface. Figure 3.13 shows the spectral reflectance of metal surfaces at normal incidence angle.

3.4.4 Fresnel reflectance

The Fresnel equation for reflection from a dielectric or metal surface depends on the polarization of the incident light relative to the plane of the surface.^{20,22} For polarized light perpendicular to the surface, the reflectance from a single surface is given by

$$\rho_{\perp} = \left(\frac{n_i \cos \theta_i - n_t \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t} \right)^2, \quad (3.28)$$

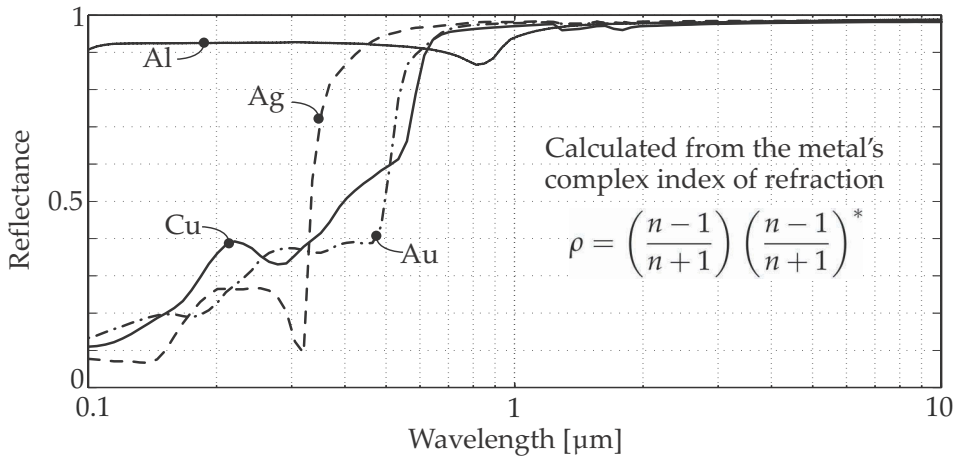


Figure 3.13 Metal spectral reflection at normal incidence angle.²¹

where θ_i is the angle of incidence of the light ray (relative to the surface normal vector), n_i is the refractive index of the medium hosting the incident ray, θ_t is the angle between the refracted ray and the surface normal vector, and n_t is the refractive index of the dielectric or metal.

The reflectance of polarized light parallel to the surface is given by

$$\rho_{\parallel} = \left(\frac{n_t \cos \theta_i - n_i \cos \theta_t}{n_t \cos \theta_i + n_i \cos \theta_t} \right)^2. \quad (3.29)$$

The reflectance of unpolarized light is given by

$$\rho = \frac{\rho_{\parallel} + \rho_{\perp}}{2}. \quad (3.30)$$

The relationship between incident angle and refracted angle is given by Snell's law (see Section 5.5.8),

$$n_i \sin \theta_i = n_t \sin \theta_t, \quad (3.31)$$

and the transmittance through the surface is given by $\tau = 1 - \rho$ because there is no absorption in the surface itself. Figure 3.14 shows the angular variation of Fresnel reflectance for a number of dielectric and metallic materials.

The Fresnel reflection equations, Equations (3.28) and (3.29), provide the reflection from a single surface, such as from an opaque surface. Transparent dielectric media, such as a plate of glass with two smooth surfaces, will reflect on both surfaces (see the figure in Problem 3.9). Assuming the same medium on both sides of the dielectric plate (i.e., air), the second surface reflects the same as the first surface. The medium's reflectance can be calculated by accounting for the successive reflectance by each surface, as well as the transmittance through the medium.

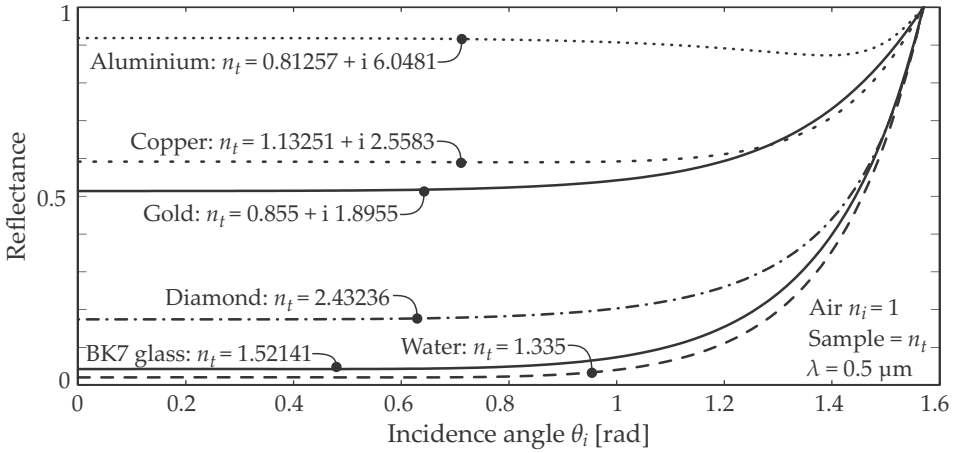


Figure 3.14 Fresnel reflection as function of incidence angle (single surface in air).²¹

3.4.5 Bidirectional reflection distribution function

The reflection from surfaces with $0 < \sigma/\lambda < 1$ does not follow the geometric laws of reflection; a more-complex function is required. The BRDF^{11,23} defines how light is reflected at an irregular or rough surface, such as shown in Figures 3.11(b) and (c). BRDF varies with wavelength — in the following discussion the monochromatic BRDF at a single wavelength is considered.

BRDF is defined as the ratio of reflected radiance $L_{\hat{S}}(d\omega_s)$ in a small solid angle $d\omega_s$ along a view vector \hat{S} to the incident irradiance $E_{\hat{I}}(d\omega_i)$ in a small solid angle $d\omega_i$ along the incidence vector \hat{I} . Note that an infinitesimally small solid angle $d\omega_i$ is considered and furthermore that the source surface with radiance $L_{\hat{I}}$ uniformly fills $d\omega_i$; hence the source surface orientation is irrelevant, and $L_{\hat{I}}d\Omega_i = L_{\hat{I}}d\omega_i$.

As shown in Figure 3.12, the direction of each of the two small solid angles is defined by the respective azimuth angles φ_i and φ_s and the zenith angles θ_i and θ_s . BRDF is therefore a four-dimensional function. Defined as L/E , BRDF has units of $[1/\text{sr}]$:

$$\text{BRDF} = f_r(d\omega_i \rightarrow d\omega_s) = \frac{L_{\hat{S}}(d\omega_s)}{E_{\hat{I}}(d\omega_i)} = \frac{L_{\hat{S}}(d\omega_s)}{L_{\hat{I}}(d\omega_i) \cos \theta_i d\omega_i}. \quad (3.32)$$

Some BRDFs are isotropic when rotated around the normal \hat{N} , yielding a three-dimensional function $f_r(\theta_i, \theta_s, \varphi_i - \varphi_s)$, whereas others are anisotropic.

The theoretical requirements for the BRDF function include positivity:

$$f_r(\theta_i, \theta_s, \varphi_i, \varphi_s) \geq 0; \quad (3.33)$$

it must conserve energy:

$$\int_{\Omega} f_r(\theta_i, \theta_s, \varphi_i, \varphi_s) \cos(\theta_s) d\omega_s \leq 1; \quad (3.34)$$

and it must be reciprocal:

$$f_r(d\omega_i \rightarrow d\omega_s) = f_r(d\omega_s \rightarrow d\omega_i). \quad (3.35)$$

A physical BRDF may not be expressible in mathematical format, thus a number of computationally viable approximations have been proposed. Often, these approximations do not comply with the theoretical requirements for the BRDF statement (e.g., conservation of energy or reciprocity). The form of such approximations depends on the domain of application and the properties of the surface. This surface could be the surface of a single object, or it could be an aggregation of a number of smaller objects, e.g., a forest tree canopy comprising many trees. Even for a single object, the surface roughness could be an aggregation of smaller peaks and valleys. The spatial structure in the surface roughness gives rise to the shape of the BRDF function.

At the one extreme, the mathematically simplest BRDF is diffuse Lambertian reflection, where $f_r(d\omega_i \rightarrow d\omega_s) = k_d$ is a constant value. Assuming that the BRDF reflects a fraction ρ of the incident light, then by the conservation of energy property [and following a derivation similar to Equation (2.10)]:

$$\begin{aligned} \rho &= \int_{\Omega} f_r(\theta_i, \theta_s, \varphi_i, \varphi_s) \cos(\theta_s) d\omega_s \\ &= k_d \int_0^{2\pi} \int_0^{\pi/2} \cos(\theta_s) d\theta_s \sin \theta_s d\varphi_s \\ &= k_d \pi, \end{aligned} \quad (3.36)$$

and hence $f_{r,\text{Lambertian}}(d\omega_i \rightarrow d\omega_s) = \rho/\pi$.

At the other extreme, the mirror reflection of a light ray \hat{I} changes the propagation direction to \hat{R} (with no light lost in any other direction). The BRDF is zero for all angles, except at $\theta_s = \theta_i$ and $\varphi_s = \varphi_i \pm \pi$. By conservation of energy, it can be shown that the BRDF for a perfect mirror is

$$f_{r,\text{Mirror}}(d\omega_i \rightarrow d\omega_s) = \frac{\rho_s(\theta_i) \delta(\theta_i - \theta_s) \delta(\varphi_i - \varphi_s \pm \pi)}{\cos \theta_i}, \quad (3.37)$$

where $\delta()$ is the Dirac delta function, and $\rho_s(\theta_i)$ is the specular surface reflectance at the angle θ_i . The presence of the $\cos \theta_i$ factor deserves further

consideration. From Equations (3.32) and (3.37) the reflected radiance from a mirror is

$$\begin{aligned} L_{\hat{S}}(d\omega_s) &= \frac{\rho_s(\theta_i)\delta(\theta_i - \theta_s)\delta(\varphi_i - \varphi_s \pm \pi)}{\cos \theta_i} L_{\hat{I}}(d\omega_i) \cos \theta_i d\omega_i \\ &= \rho_s(\theta_i) L_{\hat{I}}(d\omega_i). \end{aligned} \quad (3.38)$$

Most real-world surfaces are neither Lambertian nor perfect mirrors. It is almost impossible to find analytic formulations for the BRDF of surfaces with complex micro-scale qualities. Three approaches are used for modeling BRDF: (1) theoretical models, attempting to model underlying processes from first principles; (2) descriptive models, attempting to fit analytical curves to trends in measured data; and (3) data-driven lookup tables based on ensemble averages of measured data.

BRDF descriptions for the visual and near-infrared spectrum include the descriptive models by Phong²⁴ (fit of a cosine^{*n*} shape), Blinn–Phong,²⁵ Ward²⁶ (fit of a Gaussian shape), Lafortune,²⁷ and Ashikhmin.²⁸ Theoretical models include the work by Cook–Torrance,^{29,30} Torrance–Sparrow,³¹ He,³² and Oren and Nayar.³³

Models developed or used³⁴ for the infrared spectrum include descriptive models by Conant and LeCompt,³⁵ Ashikhmin,²⁸ and Sandford–Robertson.³⁶ Theoretical models include models by Priest–Germer,³⁷ Cailault,³⁸ Beard–Maxwell,³⁹ Cook–Torrance,^{29,30} and Snyder and Wan.¹⁹

The relatively simple Phong phenomenological BRDF model that includes diffuse and specular components is

$$f_{r,\text{Phong}} = \frac{\rho_d}{\pi} + \frac{\rho_s(n+1)\cos^n \alpha}{2\pi \cos \theta_i}, \quad (3.39)$$

where the angles are defined in Figure 3.12, ρ_d is the diffuse reflection constant, n determines the angular divergence of the lobe, and ρ_s determines the peak value or ‘strength’ of the lobe. Energy conservation requires that $\rho_s + \rho_d = \rho$, where $\rho = \Phi_r/\Phi_i$ is the total reflected flux divided by the total incident flux. The Phong model is not a mathematically compliant BRDF because for large α the BRDF value could be negative (i.e., the lobe enters below the surface) — and if the BRDF value is set to zero for such cases, the law of energy conservation is violated. The Phong model therefore fails at large α angles and for small n . The Phong model also does not comply with the requirement for reciprocity. There are several variations on the Phong theme that attempt to achieve increased accuracy. Four typical Phong specular reflection profiles are shown in Figure 3.15.

The Cook–Torrance BRDF model²⁹ considers the surface to consist of a large number of micro-facets, with mirror reflections off each facet. The

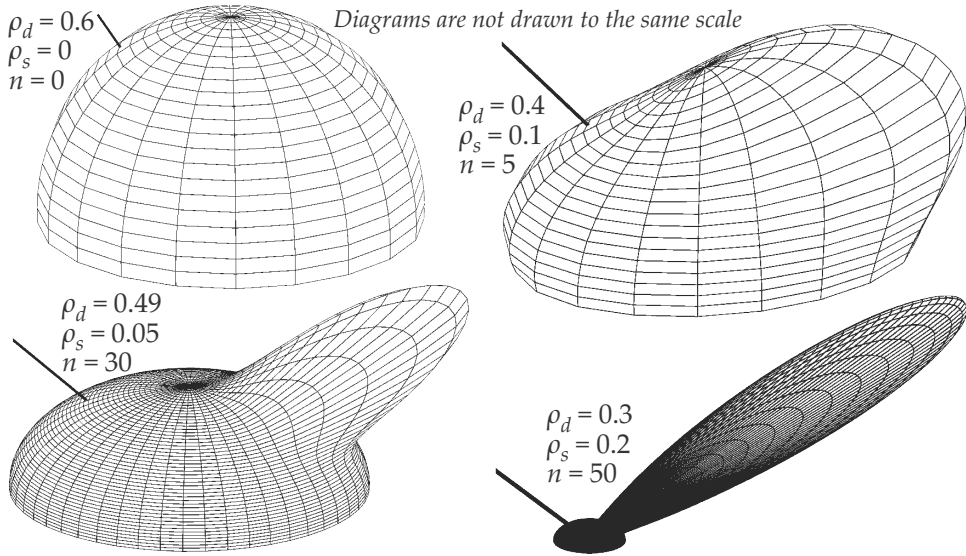


Figure 3.15 BRDF calculated by the simple Phong equation.

theoretically derived model is too involved for detailed coverage here. In summary, the BRDF is given by

$$f_{r,\text{Cook-Torrance}} = \frac{\rho_d}{\pi} + \frac{\rho_s F_\lambda D G}{\pi (\hat{N} \cdot \hat{S})(\hat{N} \cdot \hat{I})}, \quad (3.40)$$

where $\rho_d + \rho_s = \rho$, D is the distribution function of micro-facet orientations, G is a geometrical attenuation factor to account for masking and shadowing, and $F_\lambda \approx [1 + (\hat{S} \cdot \hat{N})]^\lambda$ models the reflection for each micro-facet. Physics-based models such as the Cook-Torrance are more accurate but also require much more extensive modeling and run-time calculation.

A series of BRDF measurements¹⁸ were made available on the Internet.⁴⁰ Figure 3.16 shows a few samples from the database. Comparison with Figure 3.15 indicates that the Phong model is limited in its ability to model real materials. In particular, the ‘Red fabric 2’ sample has a significant amount of back reflection, which the Phong model cannot provide.

3.5 Directional Emissivity

The preceding section clearly demonstrates surface directional reflectance properties. By the same physical mechanisms, emissivity can also have directional properties. For an opaque surface the conservation of flux requires that $\epsilon(\theta_s, \varphi_s) = 1 - \rho(\theta_s, \varphi_s)$. Therefore, much that applies to directional reflectance also applies to directional emissivity.

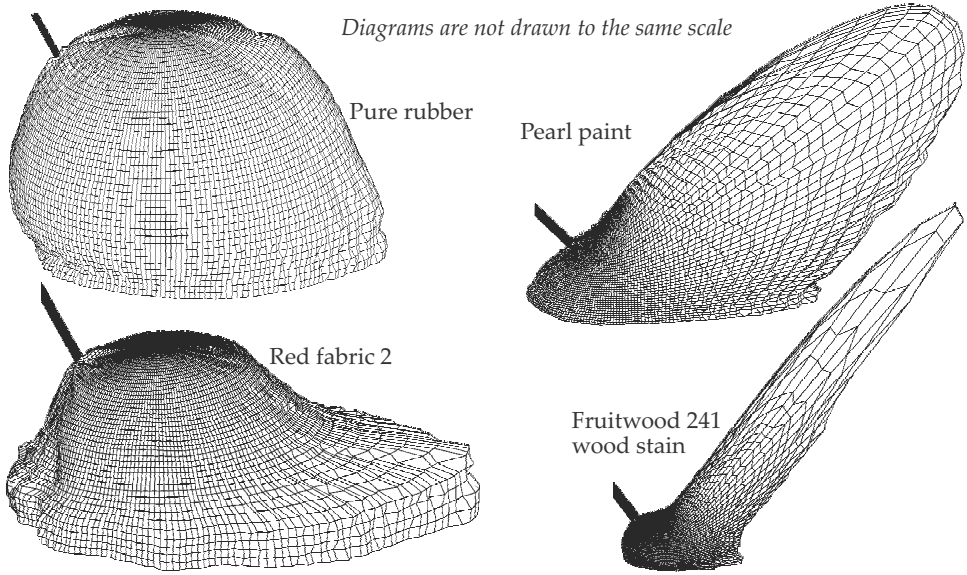


Figure 3.16 BRDF measured in the visual spectrum.¹⁸

Ignoring the incidence and reflection vectors \hat{I} and \hat{R} in Figure 3.12, the spectral directional emissivity of a thermal radiator along vector \hat{S} is given by

$$\epsilon_{\lambda}(\theta_S, \varphi_S) = \frac{L_{\lambda}(T, \theta_S, \varphi_S)}{L_{\text{bb}\lambda}(T)}, \quad (3.41)$$

where $L_{\lambda}(T, \theta_S, \varphi_S)$ is the source radiance, and $L_{\text{bb}\lambda}(T)$ is the blackbody radiation at the same temperature T .

The definition of the spectral hemispherical emissivity is

$$\epsilon_{\lambda} = \frac{M_{\lambda}(T)}{M_{\text{bb}\lambda}(T)} = \frac{\int_0^{2\pi} \int_0^{\pi/2} L_{\lambda S}(\theta_S, \varphi_S, T) \cos \theta_S \sin \theta_S d\theta_S d\varphi_S}{\int_0^{2\pi} \int_0^{\pi/2} L_{\text{bb}\lambda}(\theta_S, \varphi_S, T) \cos \theta_S \sin \theta_S d\theta_S d\varphi_S}, \quad (3.42)$$

and by using Equation (3.41) it follows that the spectral hemispherical emissivity is given by

$$\epsilon_{\lambda} = \frac{\int_0^{2\pi} \int_0^{\pi/2} \epsilon_{\lambda}(\theta_S, \varphi_S) \cos \theta_S \sin \theta_S d\theta_S d\varphi_S}{\int_0^{2\pi} \int_0^{\pi/2} \cos \theta_S \sin \theta_S d\theta_S d\varphi_S} \quad (3.43)$$

$$= \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} \epsilon_{\lambda}(\theta_S, \varphi_S) \cos \theta_S \sin \theta_S d\theta_S d\varphi_S. \quad (3.44)$$

The directional emissivity of a smooth, metallic surface can be calculated from the Fresnel reflectance equations, using $\tau + \rho + \epsilon = 1$ and

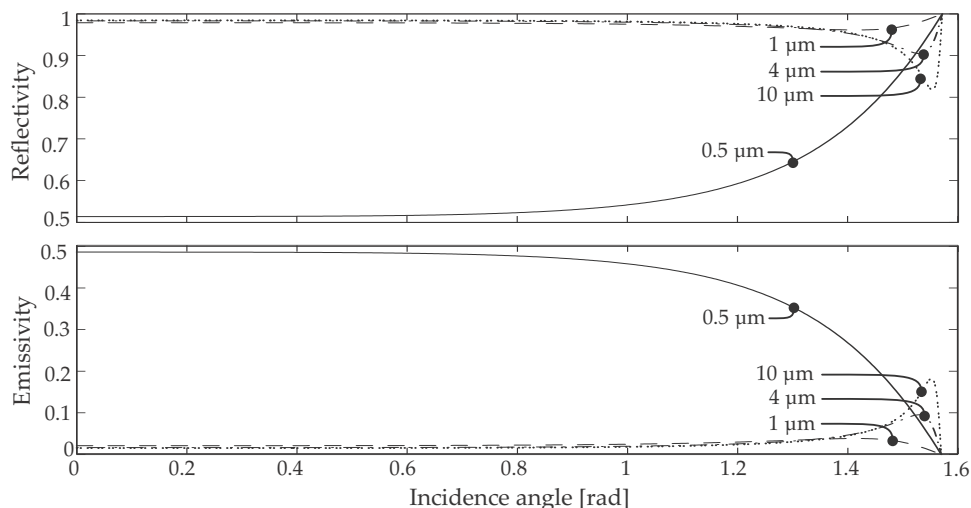


Figure 3.17 Directional reflectance and emissivity for a smooth, gold surface.

noting that $\tau = 0$ except for very thin layers. Figure 3.17 shows the directional reflectivity and emissivity of a smooth, gold surface, calculated from refractive index data.²¹ It is evident that the surface emissivity is approximately constant for incidence angles up to 0.8 rad, but the angular variation at larger incidence angles could increase (at longer wavelengths) but eventually decrease to zero at zero-grazing angles. Reviewing the Fresnel reflectance for dielectrics in Figure 3.14, it is evident that the dielectric surface emissivity ($\epsilon = 1 - \rho$) stays approximately constant up to incidence angles of 1 rad, whereafter it decreases to zero for a $\pi/3$ rad incidence angle.

Dielectrics and metals both exhibit directional emissivity, as shown above. It is evident from Figure 3.17 that at longer infrared wavelengths, the hemispherical emissivity will not differ significantly from the directional emissivity at normal incidence angle.⁴¹ The ratio of hemispherical emissivity to directional emissivity $r = \epsilon/\epsilon(0,0)$ for metals is rarely outside $1 \leq r \leq 1.3$ except at large incidence angles (see Figure 3.17). For nonconductor dielectrics the ratio is generally $0.95 \leq r \leq 1$. It is clear from Figure 3.17 that this generalization does not hold in the visual spectral band. Note that for a Lambertian radiator $r = 1$.

3.6 Directional Reflectance and Emissivity in Nature

In Section 3.4 it is shown how (random) surface roughness affects the surface reflectance. Surfaces with *directionally structured* roughness may also exhibit directional reflectance and emissivity. An example of such a struc-

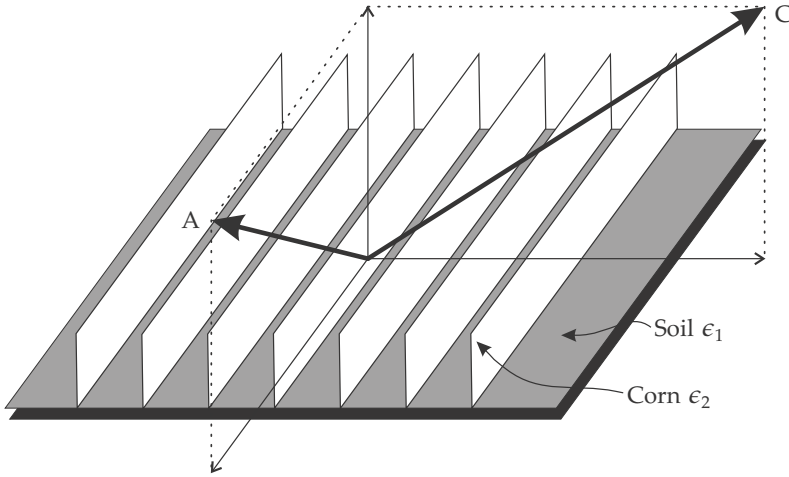


Figure 3.18 Structured scene content in a corn field.

ture is a field with rows of corn, shown in Figure 3.18. The soil has emissivity ϵ_1 , whereas the corn has emissivity ϵ_2 . In the along-row direction (A), the projected area of the corn is small, and the observer sees mostly ground. In the cross-row direction (C), at low-elevation angles, only the corn is visible and not the ground. Directional emissivity is observed in environments such as crop lands, snow, ground quartz sand, i.e., any surface with one or more materials in ordered spatial structure.

3.7 The Sun

The sun plays a major role in optical signatures by surface reflectance at shorter wavelengths, and increased self-exittance at longer wavelengths because of increased surface temperature. In this section a simple model for reflected sun radiance is derived. The distance between the sun and the earth is approximately $R_{\text{sun}} = 149 \times 10^6$ km. The sun's diameter is approximately 1.39×10^6 km. Simple calculation shows that the sun's angular size, subtended from the earth, is approximately 0.534 degrees (68.3 μsr). The sun's surface radiance can be modeled by a thermal radiator with a temperature T_s of 5800 K to 5900 K, even though the temperature inside the sun is much higher.

The solar irradiance on an object on the earth's surface is given by

$$E_{\lambda\text{sun}} = \frac{\epsilon_s L_{\text{bb}\lambda}(T_s) A_{\text{sun}} \tau_s \cos \theta_i}{R_{\text{sun}}^2}, \quad (3.45)$$

where $\epsilon_s \approx 1$ is the emissivity of the sun's surface, T_s is the sun's surface temperature, τ_s is the atmospheric transmittance between the sun and

the surface, and θ_i is the angle between the surface normal and the sun vector. The reflected sun radiance from a perfectly Lambertian surface is then given by $L_\lambda = \rho_d E_{\lambda\text{sun}}/\pi$, where ρ_d is the surface diffuse reflectance function. The reflected radiance is then given by

$$L_\lambda = \frac{\epsilon_s L_{\text{bb}\lambda}(T_s) A_s \tau_s \rho_d \cos \theta_i}{\pi R_{\text{sun}}^2} \quad (3.46)$$

$$= \psi \epsilon_s L_{\text{bb}\lambda}(T_s) \tau_s \rho_d \cos \theta_i, \quad (3.47)$$

where $\psi = A_{\text{sun}}/(\pi R_{\text{sun}}^2) = 2.1757 \times 10^{-5}$ [sr/sr] follows from the geometry. The sun geometry factor is an inverse form of the view factor described in Section 2.8.

Bibliography

- [1] Dereniak, E. L. and Boreman, G. D., *Infrared Detectors and Systems*, John Wiley & Sons, New York (1996).
- [2] Boyd, R. W., *Radiometry and the Detection of Optical Radiation*, John Wiley & Sons, New York (1983).
- [3] Wikipedia, "Planck's Law," http://en.wikipedia.org/wiki/Planck's_law.
- [4] Toyozawa, Y., *Optical Processes in Solids*, Cambridge University Press, Cambridge, UK (2003).
- [5] Wolfe, W. L. and Zissis, G., *The Infrared Handbook*, Office of Naval Research, US Navy, Infrared Information and Analysis Center, Environmental Research Institute of Michigan (1978).
- [6] SpectralCalc, GATS Inc., "Radiance: Integrating the Planck Equation," http://www.spectralcalc.com/blackbody/integrate_planck.html.
- [7] Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series and Products*, Academic Press, New York (1981).
- [8] Mohr, P. J., Taylor, B. N., and Newell, D. B., "CODATA recommended values of the fundamental physical constants: 2010," *Rev. Mod. Phys.* 84(4), 1527–1605 (2012) [doi: 10.1103/RevModPhys.84.1527].
- [9] SciPy, "SciPy Reference Guide: Constants (scipy.constants)," <http://docs.scipy.org/doc/scipy/reference/constants.html#codata2010>.
- [10] Pyradi team, "Pyradi Radiometry Python Toolkit," <http://code.google.com/p/pyradi>.

- [11] Palmer, J. M. and Grant, B. G., *The Art of Radiometry*, SPIE Press, Bellingham, WA (2009) [doi: 10.1117/3.798237].
- [12] Wyatt, C. L., *Radiometric Calibration: Theory and Methods*, Academic Press, New York (1978).
- [13] Lienhard IV, J. H. and Lienhard V, J. H., *A Heat Transfer Textbook*, Phlogiston Press, Cambridge, MA (2003).
- [14] Spectral Sciences Inc. and U. S. Air Force Research Laboratory, "MODTRAN," modtran5.com.
- [15] Wen, C. and Mudawar, I., "Modeling the effects of surface roughness on the emissivity of aluminum alloys," *International Journal of Heat and Mass Transfer* 49, 4279–4289 (2006).
- [16] Black, W. Z. and Schoenhals, R. J., "A study of directional radiation properties of specially prepared V-groove cavities," *Journal of Heat Transfer* 90, 420–428 (1968).
- [17] Nicodemus, F. E., "Normalization in Radiometry," *Applied Optics* 12, 2960–2973 (1973).
- [18] Matusik, W., Pfister, H., Brand, M., and McMillan, L., "A Data-Driven Reflectance Model," *ACM Transactions on Graphics* 22(3), 759–769 (July 2003).
- [19] Snyder, W. C. and Wan, Z., "BRDF Models to Predict Spectral Reflectance and Emissivity in the Thermal Infrared," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* 36, 214–225 (1998).
- [20] Hecht, E., *Optics*, 4th Ed., Addison Wesley, Boston, MA (2002).
- [21] Polyanskiy, M., "RefractiveIndex Info," <http://refractiveindex.info/>.
- [22] Born, M. and Wolf, E., *Principles of Optics*, 7th Ed., Pergamon Press, Oxford, UK (2000).
- [23] Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsburg, I. W., and Limperis, T., "Geometrical considerations and nomenclature for reflectance," NBS monograph 160, National Bureau of Standards (October 1977).
- [24] Phong, B. T., "Illumination for computer generated pictures," *Communications of ACM* 6, 311–317 (1975).

- [25] Blinn, J. F., "Models of light reflection for computer synthesized pictures," *4th Annual Conference on Computer Graphics and Interactive Techniques* 192 (1977) [doi: 10.1145/563858.563893].
- [26] Ward, G. J., "Measuring and modeling anisotropic reflection," *SIGGRAPH 92* (1992) [doi: 10.1145/133994.134078.].
- [27] Lafortune, E., Foo, S., Torrance, K., and Greenberg, D., "Non-linear approximation of reflectance functions," *SIGGRAPH 97* (1997).
- [28] Ashikhmin, M. and Shirley, P., "An Anisotropic Phong BRDF Model," *Journal of Graphics Tools* 5, 25–32 (2000).
- [29] Cook, R. and Torrance, K., "A reflectance model for computer graphics," *SIGGRAPH* 15, 301–316 (1981).
- [30] Cook, R. L. and Torrance, K. E., "A Reflectance Model for Computer Graphics," *ACM Transactions on Graphics* 1, 1, 7–24 (January 1982).
- [31] Torrance, K. and Sparrow, E., "Theory for Off-Specular Reflection from Roughened Surfaces," *J. Optical Soc. America* 57, 1105–1114 (1976).
- [32] He, X., Torrance, K., Sillion, F., and Greenberg, D., "A comprehensive physical model for light reflection," *Computer Graphics* 25, 175–186 (1991).
- [33] Nayar, S. and Oren, M., "Generalization of the Lambertian Model and Implications for Machine Vision," *International Journal on Computer Vision* 14, 227–251 (1995).
- [34] Brady, A. and Kharabash, S., "Further Studies into Synthetic Image Generation using CameoSim," Tech. Rep. DSTO-TR-2589, Intelligence, Surveillance and Reconnaissance Division Defence Science and Technology Organisation (2011).
- [35] Dudzik, M. C., Ed., *The Infrared and Electro-Optical Systems Handbook: Electro-Optical System Design, Analysis and Testing*, Vol. 4, ERIM and SPIE Press, Bellingham, WA (1993).
- [36] Sandford, B. P. and Robertson, D. C., "Infrared Reflectance Properties of Aircraft Paints," tech. rep., Philips Laboratory, Geophysics Directorate/G-POA (August 1994).
- [37] Priest, R. G. and Germer, T. A., "Polarimetric BRDF in the Microfacet Model: Theory and Measurements," *2000 Meeting of the Military Sensing Symposia Specialty Group on Passive Sensors* (2002).

- [38] Caillault, K., Fauqueux, S., Bourlier, C., and Simoneau, P., "Infrared multiscale sea surface modeling," *Proc. SPIE* **6360**, 636006 (2006) [doi: 10.1117/12.689720].
- [39] Maxwell, J., Beard, J., Weiner, S., and Ladd, D., "Bidirectional reflectance model validation and utilization," Tech. Rep. AFAL-TR-73-303, Research Institute of Michigan (ERIM) (October 1973).
- [40] Matusik, W., "MERL BRDF Database," <http://www.merl.com/brdf/>.
- [41] Incropera, F. P., De Witt, D. P., Bergman, T. L., and Lavine, A. S., *Fundamentals of Heat and Mass Transfer*, 6th Ed., John Wiley & Sons, New York (2007).
- [42] Colors of heated steel, http://www.sizes.com/materlsls/colors_of_heated_metals.htm.
- [43] Wikipedia, "Planckian locus," http://en.wikipedia.org/wiki/Planckian_locus.
- [44] Wikipedia, "CIE 1931 color space," http://en.wikipedia.org/wiki/CIE_1931_color_space.
- [45] Kirk, R., "Standard Colour Spaces," Technical Note FL-TL-TN-0139-StdColourSpaces, FilmLight Digial Film Technology (2007).
- [46] Wikipedia, "Standard illuminant," http://en.wikipedia.org/wiki/Standard_illuminant.
- [47] Wikipedia, "White point," http://en.wikipedia.org/wiki/White_point.
- [48] Pyradi team, "Pyradi data," <https://code.google.com/p/pyradi/source/browse>.
- [49] Her, T., Finlay, R. J., Wu, C., Deliwala, S., and Mazur, E., "Microstructuring of silicon with femtosecond laser pulses," *Applied Physics Letters* **73**, 1673–1675 (1998).
- [50] Brown, R. J. C., Brewer, P. J., and Milton, M. J. T., "The physical and chemical properties of electroless nickel-phosphorus alloys and low reflectance nickel-phosphorus black surfaces," *Journal of Materials Chemistry* **12**, 2749–2754 (2002).
- [51] Xi, J., Schubert, M. F., , and E. Fred Schubert, J. K. K., Chen, M., Lin, S.-Y., Liu, W., and Smart, J. A., "Optical thin-film materials with low refractive index for broadband elimination of Fresnel reflection," *Nature Photonics* **1**, 176 – 179 (2007) [doi: doi:10.1038/nphoton.2007.26].

- [52] Tsakalakos, L., Balch, J., Fronheiser, J., Shih, M.-Y., LeBoeuf, S. F., Pietrzykowski, M., Codella, P. J., Korevaar, B. A., Sulima, O., Rand, J., Davuluru, A., , and Rapol, U., "Strong broadband optical absorption in silicon nanowire films," *Journal of Nanophotonics* 1 (2007) [doi: DOI: 10.1117/1.2768999].
- [53] University of California, S. B., "MODIS UCSB Emissivity Library," <http://www.icesb.ucsb.edu/modis/EMIS/html/em.html>.

Problems

- 3.1 Calculate the color coordinates and show the approximate colors for blackbodies at the following temperatures: 3200 K, 5000 K, 6500 K, and 9000 K. [4]
- 3.2 Calculate and plot graphically (more than 10 data points) the radiance from the surface of a Planck thermal radiator at a temperature of 1000 K, over the spectral range of 3–5 μm , as follows: (1) in the wavelength domain (in units of $[\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \mu\text{m})]$), (2) in the wavenumber domain (in units of $[\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \text{cm}^{-1})]$), and (3) convert between the results obtained in (1) and (2) above, using the conversion defined in Section 2.3.3. [6]
- 3.3 Calculate the amount of heat energy in joules flowing into a beef steak on a outdoor barbecue grid from the moment it is put onto the grid, until it is ready to eat. You may consult any source (except fellow students) but provide the reference to the source. Clearly state and motivate all assumptions. Apply the golden rules to the problem, i.e., dimensional analysis, developing a good mathematical definition, and drawing detailed diagrams of the problem statement. Include all code or numerical files used in the calculation. [6]
- 3.4 Ironsmiths use the color of a steel sample to estimate the temperature of the sample. Subjective descriptions, such as 'dull cherry-red,' are used to describe the temperature, as in the following table:⁴²

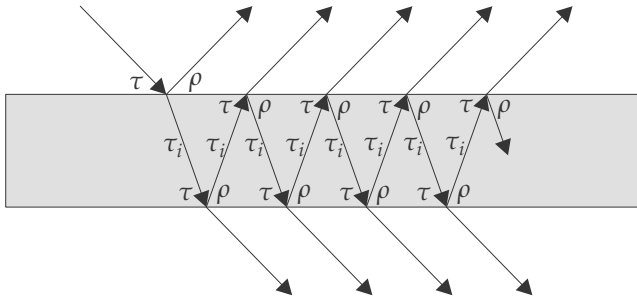
Temperature judged by color			
Color	Halcomb	Howe	White
	°C	°C	°C
Red heat, visible in the dark	400	470	.
Red heat, visible in the twilight	474		.
Red heat, visible in the daylight	525	475	532
Red heat, visible in the sunlight	581		556
Dark red	700	550–625	635
Dull cherry-red	800		677
Cherry-red	900	700	746
Bright cherry-red	1000	850	843
Orange-red	1100		899
Orange-yellow	1200	950–1000	941
Yellow-white	1300	1050	996
White welding heat	1400	1150	1079
Brilliant white	1500		1205
Dazzling white (bluish-white)	1600	.	.

The objective with this investigation is to confirm the information in this table and to seek a more ‘scientific’ manner to link color to temperature.

- 3.4.1 Use the given color-matching function data to calculate the xy chromaticity color coordinates for a thermal radiator (the Planckian locus), with the following temperatures: $T \in \{500, 1000, 1500, 2000, 2500, 3000, 4000, 6000, 10000, 1 \times 10^{10}\}$ K. [4]
- 3.4.2 Confirm your calculated color coordinates against the values shown in Figure A.1. Consult a properly colored chromaticity diagram^{43,44} and comment on the ironsmith color table given above. [4]
- 3.4.3 Elaborate on the color referred to as the ‘white point.’^{45–47} How stable is the color white? Give examples of acceptable white colors. On the grounds of your research, comment on the eye’s value as a scientific color instrument. [2]
- 3.5 Refer to Section 2.10.5 and repeat the calculations there with your own code. The data is available on the pyradi website.⁴⁸
- 3.5.1 Calculate the color coordinates of the six samples when illuminated by the four source spectral radiances. Calculate and plot your own versions of the graphs in Figures 2.17 and 2.18. Enter the color coordinate values into the table below and comment on your observations. Apply the Golden Rules (Chapter 10) in the derivation of the solution. [9]

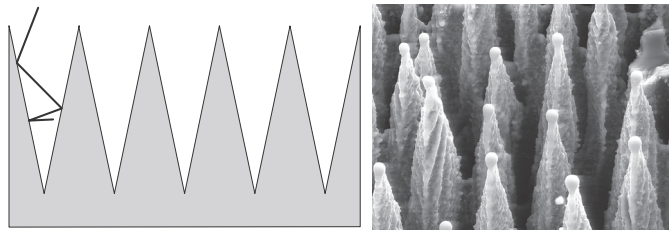
	Tomato	Lettuce	Prune	Leaf	Glove	Paper
Fluorescent						
Sunlight						
Incandescent						
Sodium lamp						

- 3.5.2 Plot the color coordinate locus for each sample when illuminated by the different sources onto the color CIE diagram (Figure 2.19). [4]
- 3.5.3 Explain why the three samples' color coordinates cluster so closely together when observed under the sodium lamp. Near which monochromatic wavelength does this cluster occur? Why? When buying paint or clothes, under which source illumination can you best compare colors? [2]
- 3.6 Calculate the total flux flowing between two circular disks with diameter 1 m, separated by 10 m. The first object has a temperature of 450 K and an emissivity of 0.5, and the second object has a temperature of 450 K and an emissivity ranging from 0 to 1.0 in steps of 0.1. Derive a mathematical equation for the net flux flow and then plot the values. [6]
- 3.7 Calculate the total flux flowing between two circular disks with diameter 1 m, separated by 10 m. The first object has a temperature of 450 K and an emissivity of 0.5, and the second object has an emissivity of 0.5 and a temperature ranging from 300 K to 600 K in steps of 50 K. Derive a mathematical equation for the net flux flow and then plot the values. [6]
- 3.8 An opaque object has a diffuse reflectance of 1.0 for wavelengths from 0 to 0.5 μm ; and 0 for wavelengths from 0.5 μm to infinity. The object is illuminated by a source with temperatures ranging from 1000 to 4000 K in steps of 1000 K. The object's own internal temperature ranges from 1000 to 4000 K in steps of 1000 K. Assume Planck law radiation in both cases. Calculate and plot the object's color coordinates for all of the temperature combinations. Validate your results against Figure A.1. Comment on your observations. [10]
- 3.9 In a thin, transparent dielectric medium, the incident flux is reflected between the medium's surfaces, as shown below. The medium itself has an internal transmittance τ_i .



- 3.9.1 Show that the reflection and transmittance magnitudes of the top and bottom surfaces are the same. [4]
- 3.9.2 Derive an equation for the reflectance (sum of all components) and transmittance of the medium. [6]
- 3.10 Calculate the solar irradiance at the top of the atmosphere. [3]
Calculate the total solar flux absorbed by the earth. Assume the earth's diameter to be 12,756.8 km. The earth has an average albedo (reflectance) of 0.3. [5]
- 3.11 In Section 3.7 a model is developed for the reflected sunlight from an object on the earth. Confirm Equation (3.47) by applying the Golden Rule of dimensional analysis. [4]
- 3.12 Use Figure 3.4 to estimate the percentage of radiance in (1) the 3–5- μm spectral band for a jet tailpipe, and (2) the 0.4–0.7- μm visual spectral band for a tungsten lamp. [2]
Repeat, but calculate more accurately using a spectral integral of the Planck law between the two wavelengths. [4]
- 3.13 Derive an equation for reflected sun radiance, similar to Equation (3.47), but for a diffuse surface modeled with the Phong BRDF function. [2]
- 3.14 A hemispherical dome of material with emissivity ϵ_1 and temperature T_1 encloses a small component with area A_2 , emissivity ϵ_2 , and temperature T_2 (Figure 3.8). The objective with this investigation is to consider the net radiance emanating from area A_2 . Both materials are opaque with Lambertian surface properties.
- 3.14.1 Consider the enclosing area A_1 . What effect do the geometrical shape of A_1 and the radius R have on the net radiance L_2 from A_2 ? [2]
- 3.14.2 For emissivity $\epsilon_1 = 0.9$, temperature $T_1 = 300$ K and temperature $T_2 = 350$ K, calculate the radiance L_2 over all wavelengths, for emissivity ϵ_2 from 0 to 1 in steps of 0.1. [4]

- 3.15 A hole with diameter D and depth d is drilled into a metal block, as shown in Figure 3.9. Assume the drill bit point angle to be exactly 90 deg, forming a hole as shown in the figure. The surface emissivity of the block and inside the hole is 0.2.
- 3.15.1 Calculate the emissivity of a hole of $d/D = 5$, for the ray shown, at an angle θ ranging from 0 deg to 90 deg at 5-deg intervals. Ignore the Fresnel reflection effect. [4]
- 3.15.2 Calculate the emissivity of a hole of $d/D = 5$, for the ray shown, at an angle θ ranging from 0 deg to 90 deg at 5-deg intervals. Include the Fresnel reflection effect. [4]
- 3.15.3 Assuming the block's temperature to be 300 K, calculate the radiance in the hole opening for an angle θ ranging from 0 deg to 90 deg at 5-deg intervals. [2]
- 3.16 Low-reflectance (i.e., high emissivity) surfaces can be constructed by spatial structures that 'trap' the light by multiple reflections. Such light traps typically have conical shapes (figure below⁴⁹), wedge structures, porous surfaces,⁵⁰ or fiber structures.^{51,52}



- 3.16.1 Derive a mathematical description of the light path into a one-dimensional wedge light trap. Elaborate on the number of bounces and the mathematical requirements for the light wave never to escape from the trap. Consider a geometrical structure in your solution rather than a purely mathematical analysis. [4]
- 3.16.2 Design a one-dimensional wedge-shaped light trap to achieve an emissivity of 0.98 at an incidence angle of 5 deg. The available surface finish has a Lambertian emissivity of 0.5. [4]
- 3.16.3 For the above light trap, calculate the emissivity for angles from normal incidence up to 90 deg at 5-deg intervals. Angles are defined in the plane of the paper. [4]
- 3.16.4 Comment on the use of a two-dimensional wedge structure as a surface for a laboratory blackbody radiator. What are the advantages and disadvantages of such a surface in this application? [2]

- 3.17 Implement Equation (3.23) in a computer program and determine how many summation terms are required to achieve an accuracy of better than 1% from the ideal solution for the following three spectral bands $\{1\text{--}2\text{ }\mu\text{m}, 3\text{--}5\text{ }\mu\text{m}, 10\text{--}12\text{ }\mu\text{m}\}$ and a temperature of 700 K. Comment on the observed differences in the required number of terms. [6]
- 3.18 Select and plot spectral emissivity data from the MODIS UCSB Emissivity Library⁵³ for at least two materials in each of the following four classes: (1) water, ice, and snow; (2) soils and minerals; (3) vegetation; and (4) synthetic materials. Comment on the similarities and differences. Is there a significant difference between emissivity values for synthetic and natural materials? [6]
- 3.19 Define Snell's law and show mathematically why the light ray 'bends' on an interface between different refractive indices. Explain what will happen in a medium with a gradual change in refractive index. See Section 5.5.8 for a hint. [4]

Chapter 4

Optical Media

The farther reason looks, the greater is the haze in which it loses itself.
Johann Georg Hamann

4.1 Overview

In the context of the source–medium–sensor system model, the medium is everything between the source and the sensor. The optical medium affects the radiance field by flux attenuation, flux amplification (in the base of lasers), flux increase (path radiance) and refractive wave distortions (e.g., turbulence). The medium effects can be either static and temporally constant or temporally and spatially dynamic, such as in turbulent flow. In keeping with the theme of the book, this chapter investigates the effects of the atmosphere as a component in the system rather than on the physical processes in the atmosphere — these are covered elsewhere.^{1–6}

The atmospheric index of refraction varies with pressure and temperature. The atmospheric air movements result in eddy currents with varying temperature and pressure, resulting in cells of varying indices of refraction. These variations cause a number of different effects, depending on the magnitude of the variation, the physical area of the variation, the nature of the optical flux, and so forth. Some of the effects that can occur are beam steering, arrival-angle variations, scintillation (variation in signal strength), and visual mirage effects. This chapter considers only static media; turbulence is well documented elsewhere.^{7–11}

Flux transfer through a medium is modeled with the radiative transfer equation (RTE).^{8,12,13} The full RTE is complex, and considerable ‘engineering’ simplification is made here to convey concepts and principles. A comprehensive discussion is beyond the scope of this book. This chapter initially considers a trivially simple RTE for homogenous media and then adds more complexity to account for path radiance and inhomogeneous media. The atmosphere^{2,8} is an important component in most optical and

infrared systems and is briefly reviewed.

4.2 Optical Mediums

4.2.1 Lossy mediums

In its most-simple form, a lossy medium can be modeled by a spectrally varying magnitude τ , called the ‘transmittance.’ Transmittance is the ratio of the source energy reaching the receiver with the medium present to the energy reaching the receiver with no medium present. In general, medium transmittance is spectrally selective and can only be meaningfully defined and determined at a specific wavelength. It is implied throughout this section that all variables are spectrally varying.

The losses in the medium are mainly attributable to scattering and absorption processes. The degree of attenuation by a medium is described by the attenuation coefficient γ with units $[\text{m}^{-1}]$. In the derivation that follows it is assumed that the absorbed or scattered power is not re-radiated in the optical path. Also, in most cases the fractional change in radiance is linearly proportional to distance in the medium.¹⁴

Consider an arbitrary medium contained between $x = 0$ and $x = R$. The fractional change in radiance L along a thin layer of the medium dx is given by

$$\frac{dL}{L} = -\gamma dx. \quad (4.1)$$

Note that γ generally varies along the path and can be described as a function of the path variable x :

$$\frac{dL}{L} = -\gamma(x)dx. \quad (4.2)$$

Integrating this equation along the path between 0 and R finds

$$\begin{aligned} \int_{L(0)}^{L(R)} \frac{dL}{L} &= - \int_0^R \gamma(x)dx \\ \frac{L(R)}{L(0)} &= \exp \left[- \int_0^R \gamma(x)dx \right]. \end{aligned} \quad (4.3)$$

If the path through the medium is homogeneous, $\gamma(x)$ is constant, leading to Bouguer’s law:

$$\frac{L(R)}{L(0)} = \tau(R) = e^{-\gamma R}, \quad (4.4)$$

where γ is the medium's attenuation coefficient or extinction coefficient with units $[\text{m}^{-1}]$, and R is the distance through the medium in $[\text{m}]$. The attenuation coefficient is related to the imaginary component of the complex index of refraction (see Sections 3.4.3 and 5.5.8).

If the medium along the path is not homogeneous, and the variation of γ along the path is known, the attenuation coefficient can be written as a normalizing constant γ_o multiplied by a range-dependent function $f(x)$, $\gamma(x) = \gamma_o f(x)$. The function $f(x)$ represents the profile of γ along the optical path. The transmittance of the path in an inhomogeneous medium is then

$$\begin{aligned}\tau(R) &= \exp \left[-\gamma_o \int_0^R f(x) dx \right] \\ &= e^{-\gamma_o \bar{R}}.\end{aligned}\tag{4.5}$$

The integral

$$\bar{R} = \int_0^R f(x) dx \tag{4.6}$$

represents the horizontal distance in a homogeneous medium (with $\gamma = \gamma_o$) for which the attenuation is the same as the actual path of length R through the inhomogeneous path [with $\gamma = \gamma(x)$]. \bar{R} is also known as the equivalent path length, given γ_o .

The attenuation usually comprises two independent processes: scattering and absorption. Particles and molecules in the medium may scatter the photons, changing the photon's direction, and thereby 'removing' the photon from the flux. Photons may also be absorbed by particles and molecules, raising the particles' energy level but also removing the photon from the flux. The two processes are modeled by two attenuation coefficients,

$$\gamma = \sigma + \alpha, \tag{4.7}$$

where σ denotes the scattering attenuation coefficient in $[\text{m}^{-1}]$, and α denotes the absorption attenuation coefficient in $[\text{m}^{-1}]$.

4.2.2 Path radiance

An optical medium can add radiance to the optical path in addition to the source radiance already present along the path.¹ Path radiance is the phenomenon whereby the path injects optical flux along the line of sight. Two common examples of path radiance are the bright appearance of fog when irradiated by the sun from above and the blue appearance of the sky

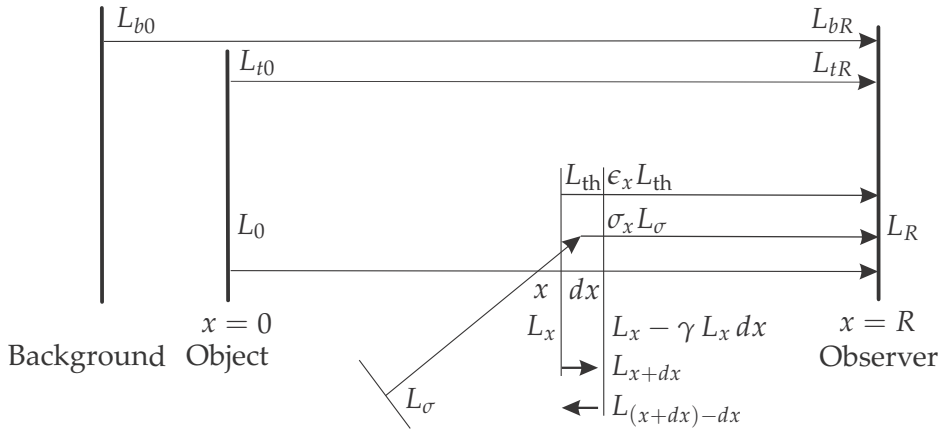


Figure 4.1 Definition of path radiance geometry variables.

on a cloudless day. In both cases sunlight is scattered into the sensor's field of view, adding flux along the line of sight. The following derivation is based on the two-flux-Kubelka–Munk theory^{15,16} developed for the optical properties of paint, but the principles apply to any medium.

Path radiance occurs from the medium's thermal self-exittance or from flux from another source that the medium scatters into the radiance field. The total flux is the sum of the source flux and all path radiance contributions.

The approach set out by Duntley¹⁷ and others^{18–20} is simplified by omitting some scattering sources, but it is extended here to include thermal self-exittance along the path. All of the variables defined here are strongly dependent upon wavelength even though it is not indicated as such. Only the conceptual model development and the results are shown here; for a detailed mathematical analysis see the sources.

Consider the path geometry shown in Figure 4.1. The 'line of sight' is defined as the direction from L_R toward the source L_0 . The optical field has a radiance L_0 at the source, and after propagating over a distance R through the medium, it has a radiance L_R . At distance x along the path the field has a radiance L_x and after propagating a further distance dx , the radiance is L_{x+dx} . Source radiance passing through dx will diminish because of attenuation in the medium, as discussed in Section 4.2.1.

The radiance L_x will increase due to the flux scattered into the line of sight by an amount $\sigma_x L_{\sigma}$, where σ_x is the scattering coefficient in $[m^{-1}]$, and L_{σ} is the external source radiance. The radiance L_x will increase furthermore by the thermal exittance of the medium by an amount $\epsilon_x L_{th}$, where ϵ_x is the medium emission coefficient with units $[m^{-1}]$, and L_{th}

is the thermal radiance of a blackbody at the same temperature as the medium. Note that this emission coefficient ϵ_x gives rise to path radiance over the distance dx in the medium, but it is not the same as emissivity as discussed in Section 3.2.

The medium attenuation and path radiance terms can be combined as a system of two interdependent differential equations for the forward and backward directions (a simplified version of the Duntley Equations¹⁷):

$$\frac{dL_{x+dx}}{dx} = -\gamma_x L_{x+dx} + \sigma_x L_\sigma + \epsilon_x L_{th} \quad (4.8)$$

and

$$-\left(\frac{dL_{(x+dx)-dx}}{dx}\right) = -\gamma_x L_x + \sigma_x L_\sigma + \epsilon_x L_{th}. \quad (4.9)$$

Assume that each variable γ_x , σ_x , and ϵ_x can be written as a normalizing constant multiplied by a common range dependent function $f(x)$, in the form $\gamma_x = \gamma_0 f(x)$, as per Equation (4.6). Now solve the simultaneous differential equations comprising Equations (4.8) and (4.9) to obtain the RTE for flux transfer together with path radiance:

$$L_R = L_0 e^{-\gamma_0 \bar{R}} + \frac{\sigma_0 L_\sigma + \epsilon_0 L_{th}}{\gamma_0} (1 - e^{-\gamma_0 \bar{R}}). \quad (4.10)$$

The solution in Equation (4.10) comprises two parts, the irradiance at the beginning of the path multiplied by the atmospheric transmittance along the path (as found in Section 4.2.1) plus the *path radiance* term. The path radiance contribution comprises the flux scattered into the line of sight and the self-emission along the line of sight. Recall from Section 2.3.4 that $1 = \alpha + \tau + \rho$; therefore, for any medium with no reflection (a reasonable assumption for a mainly transmissive medium), $\alpha = 1 - \tau$. The term $1 - \exp(-\gamma_0 \bar{R})$ now appears as an ‘emissivity’ factor multiplying the medium radiance.

By applying Kirchhoff’s law $\alpha = \epsilon$, and on condition that the scattering coefficient for attenuation and path radiance share the same value, Equation (4.7) can be written $\gamma_0 = \sigma_0 + \epsilon_0$. Equation (4.10) then becomes

$$L_R = L_0 e^{-(\sigma_0 + \alpha_0) \bar{R}} + \frac{\sigma_0 L_\sigma + \epsilon_0 L_{th}}{\sigma_0 + \epsilon_0} (1 - e^{-(\sigma_0 + \alpha_0) \bar{R}}). \quad (4.11)$$

If now $\sigma_0 = 0$, i.e., for a clear atmosphere in the far-infrared spectral region,

$$L_R = L_0 e^{-\alpha_0 \bar{R}} + L_{th} (1 - e^{-\alpha_0 \bar{R}}), \quad (4.12)$$

and it is evident that path radiance can be approximated by the medium radiance (as for any opaque surface radiator) multiplied by an atmospheric 'emissivity' (equal to unity minus the medium transmittance). This interpretation of 'emissivity' and path radiance is appropriate for the thermally radiated flux from the medium: the concepts of emissivity and radiance can be understood in physical terms. Path radiance caused by scattering is not comprehensible in terms of true emissivity and (thermal) radiance; in this case it is merely a model, not a physical reality. The conditions under which this simplification is valid are investigated in more detail later in this chapter.

4.2.3 General law of contrast reduction

Medium attenuation and path radiance reduce the apparent contrast of an object against a background radiance. When observing mountains on the horizon, one observes different mountains to have varying contrast. On a clear day, the mountain stands out clearly, and on a hazy day, the mountain disappears in the haze. This section investigates general equations describing the effects of a medium on the apparent contrast of an object in an image.

There are several different definitions of contrast. For the purpose of this section, we follow the definition¹⁷ used for visual observations. The contrast at the object ($x = 0$) and the apparent contrast at the observer ($x = R$) are given by (from the definition of contrast)

$$C_0 = \frac{L_{t0} - L_{b0}}{L_{b0}} \quad (4.13)$$

and

$$C_R = \frac{L_{tR} - L_{bR}}{L_{bR}}, \quad (4.14)$$

where the radiance terms are defined in Figure 4.1; where the footscripts denote: t as the object of interest (target), b as the background behind the object, 0 as the range at the object, and R as the range at the observer. Furthermore, from the definition of transmittance,

$$L_{tR} - L_{bR} = (L_{t0} - L_{b0})e^{-\gamma\bar{R}}. \quad (4.15)$$

Combining the last three equations, the general law of contrast reduction is obtained as

$$\tau_c = \frac{C_R}{C_0} = \left(\frac{L_{b0}}{L_{bR}}\right)e^{-\gamma\bar{R}}, \quad (4.16)$$

where L_{b0} is the background radiance at the object, and L_{bR} is the background radiance at the observer, both along the line of sight from the observer

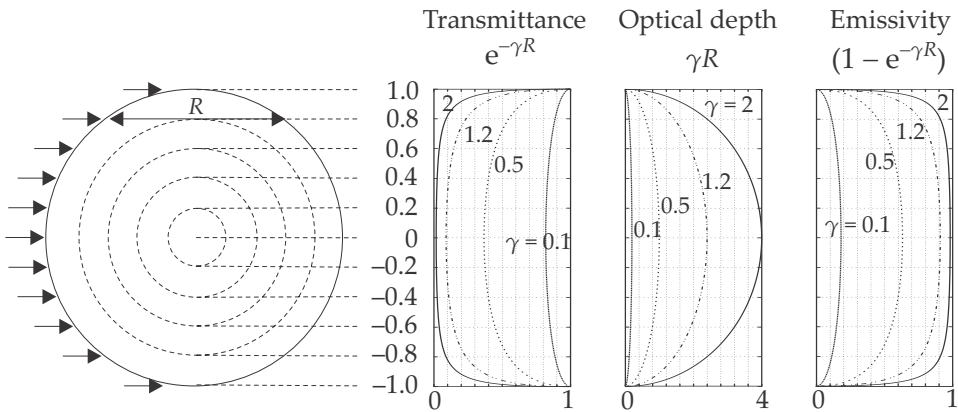


Figure 4.2 Radiance from a cylindrical hot gas plume for different values of attenuation coefficient.

to the object. The contrast transmittance is the fraction with which the atmosphere reduces the inherent (close-range) contrast in the scene. Equation (4.16) has a form similar to Bouguer's law but with a background radiance scaling modifier. Contrast transmittance in the atmosphere is considered in more detail in Section 4.6.9.

4.2.4 Optical thickness

The optical thickness $\beta = \gamma R$ of the medium is defined as the exponent in Bouguer's law. If the optical thickness is small, the transmittance is high and the medium is clear. On the other hand, if β is high, transmittance is low and the medium is opaque. It follows that optical thickness is a measure of the opacity of the medium. Optical thickness is therefore an indicator of the 'emissivity' term $(1 - e^{-\gamma R})$, as derived in the previous section.

4.2.5 Gas radiator sources

Optical thickness provides another important insight. Section 3.2.1 states that for gasses $\rho = 0$, and Section 4.2.2 shows that $\epsilon = 1 - \tau$. Hence, an optically thick gas medium with low transmittance has a high emissivity.

Consider the (path) radiance from a cylindrical, hot gas plume, as shown in Figure 4.2. The cylinder has unity radius and is filled with a gaseous medium with attenuation coefficient γ of 0.1, 0.5, 1.2, and 2. Horizontal sections through the plume have length R , which depends on the height of the section.

For $\gamma = 0.1$, the plume has a transmittance of 0.82 through the center, low optical depth, and hence a low emissivity (0.18). The total volume of this plume contributes to the radiance field from the plume. The plume acts as a *volume radiator*. This also means that the entire plume volume loses heat by thermal radiation.

At the other extreme, for $\gamma = 2$, the plume has a transmittance of 0.018 through the center, high optical depth, and hence a high emissivity (0.982). It follows that the exitance from this medium emanates from the *outer layers* of the medium. A volume of highly attenuating gas therefore acts as a *surface radiator*. The plume loses heat only from the surface; the inner volume retains its heat better than for a plume with low γ .

Conversely, an optically thin medium has high transmittance and low emissivity. Because of the low emissivity (and good transmittance), the radiation emanates from the whole volume of the medium — a *volume radiator*.

The emissivity of a radiating gas is a spectral variable, as shown in Figure 3.7. A CO₂ gas plume has high emissivity (large γ) at 4.35 μm , but almost zero emissivity at 3.5 μm . The plume can therefore be a surface radiator (opaque as brick wall) at 4.35 μm but it is fully transparent at 3.5 μm .

4.3 Inhomogeneous Media and Discrete Ordinates

Most media are not homogeneous, $f(x) \neq c$, and Equations (4.8) and (4.9) must be solved by integration along the optical path. Because the profile $f(x)$ is almost never an analytical function, the integral cannot be solved analytically. Practical solutions to the RTE are all based on discretization^{4,21} of the continuous function $f(x)$ into discrete parts $f_d[X_{di}]$. Each X_d is a discrete section along the path x , such that the error $f(x) - f_d[X_{di}]$ is sufficiently small for all values of x and corresponding values of X_{di} . By decreasing the discrete interval X_d toward zero, the difference between the continuous function and the discrete approximation approaches zero (at least sufficiently so for engineering purposes!).

Practical computational considerations limit the number of discrete intervals and hence determine the coarseness of the approximation. In some cases the discrete interval is not necessarily a constant value. One such example is the modeling of atmospheric vertical profiles as 36 discrete layers (see Section 4.7). A reasonable modeling approach would be to adapt the sampling interval X_d to the relative magnitude of the profile f . For path regions with low values of f , the interval X_d could be modified

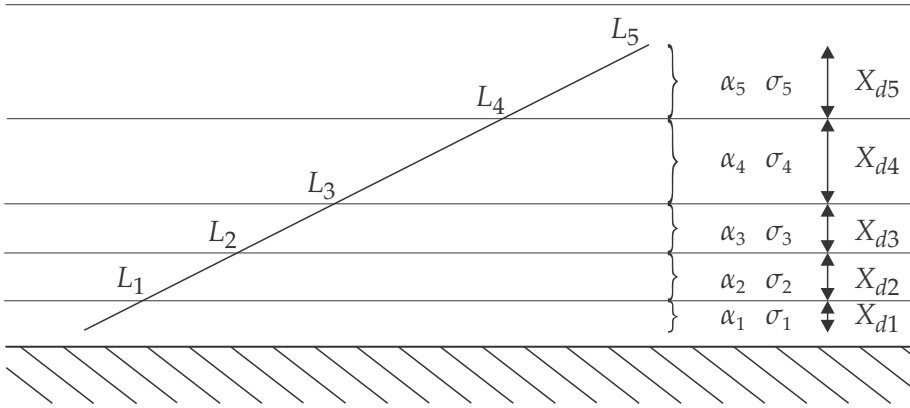


Figure 4.3 Discrete vertical strata.

such that the product $X_{di} f_d[X_{di}]$ is approximately constant for all values of i . In this strategy, each discrete path section contributes equally to the integral along the path.

The discrete model intervals X_{di} may be larger than the path length itself, and the path's two end points may end up in between X_{di} boundaries (Figure 4.3). The endpoint section contributions are added pro-rata, according to the length in each endpoint.

The geometric shape of the discrete intervals depends on the problem at hand. The earth's atmosphere on a global scale is modeled as a set of concentric shells, centered on the earth's center. On a smaller, local scale, the atmosphere can be modeled as a set of vertical strata. An aircraft plume can be modeled as a stacked set of short concentric cylinders. An arbitrary volumetric radiator can be modeled as a set of voxels, stacked like a three-dimensional chessboard. The important consideration is that within each of these individual volumes, the medium is considered uniform and homogeneous. The solution of the RTE within these discretized models require numerical computation.

4.4 Effective Transmittance

From Equations (2.26) and (4.4) (and looking ahead to Section 7.2.2) the irradiance at the sensor from a distant source is given by an equation of the form as shown in Equation (6.19),

$$E = \frac{A_0}{R^2} \int_0^\infty \epsilon_\lambda L_\lambda(T_s) \tau_{a\lambda}(R) \mathcal{S}_\lambda d\lambda, \quad (4.17)$$

where A_0 is the area of the receiver, R is the range from the source to the receiver, ϵ_λ is the source spectral emissivity, $L_\lambda(T_s)$ is the Planck law for

an object at temperature T_s , $\tau_{a\lambda}(R)$ is the medium (atmosphere) spectral transmittance at the range R , \mathcal{S}_λ is the sensor spectral response. The integral must be recalculated for each source-to-object distance because the atmospheric transmittance varies with distance. With a few simplifications, it is possible to re-write the irradiance equation in terms of the 'effective transmittance' as follows:

$$E = \frac{A_0 \tau_a(R)}{R^2} \int_0^\infty \epsilon_\lambda L_\lambda(T_s) \mathcal{S}_\lambda d\lambda, \quad (4.18)$$

where the integral is independent of range and is only calculated once. The effective transmittance $\tau_a(R)$ is then given by

$$\tau_a(R) = \frac{\int_0^\infty \epsilon_\lambda L_\lambda(T_s) \tau_{a\lambda}(R) \mathcal{S}_\lambda d\lambda}{\int_0^\infty \epsilon_\lambda L_\lambda(T_s) \mathcal{S}_\lambda d\lambda}, \quad (4.19)$$

a scalar number instead of a spectral quantity. The effective transmittance is spectrally weighted by the source and sensor spectral responses, and therefore accounts for the effect of the system's spectral quantities. The designer may be tempted to calculate the effective transmittance only once and thereafter use only the effective value, but caveat emptor!

This definition of effective transmittance is the only accurate way to describe average or effective transmittance. The MODTRANTM `tape6` 'average' transmittance should not be used because *it does not include the sensor or the source spectral properties*. MODTRANTM provides separate facilities to calculate a weighted transmittance.

Figure 4.4 shows the atmospheric spectral transmittance for several path lengths, a sensor response, and the normalized spectral radiance of several sources. The sources include a CO₂ gas radiator (aircraft plume) and thermal radiators at a range of temperatures. Equation (4.19) was used to determine the effective transmittance for the different sources. The sensor response shown in Figure 4.4 was used. The effective transmittance values so determined are shown in the bottom graph of Figure 4.4.

Note the severe attenuation for the CO₂ gas plume, even over short ranges. This stems from the fact that the CO₂ spectral exitance is severely attenuated by the cold CO₂ in the atmosphere. This effect is clearly visible in the spectral plot in Figure 4.4.

Less dramatic, but still relevant, is the variation of effective transmittance with source temperature. If the effective transmittance for a 2300-K source is used for an object at 330 K, a large range error arises for a given transmittance, especially at longer ranges. Consider the line for 0.18 effective transmittance in Figure 4.4, which gives 10 km for a 2300-K source,

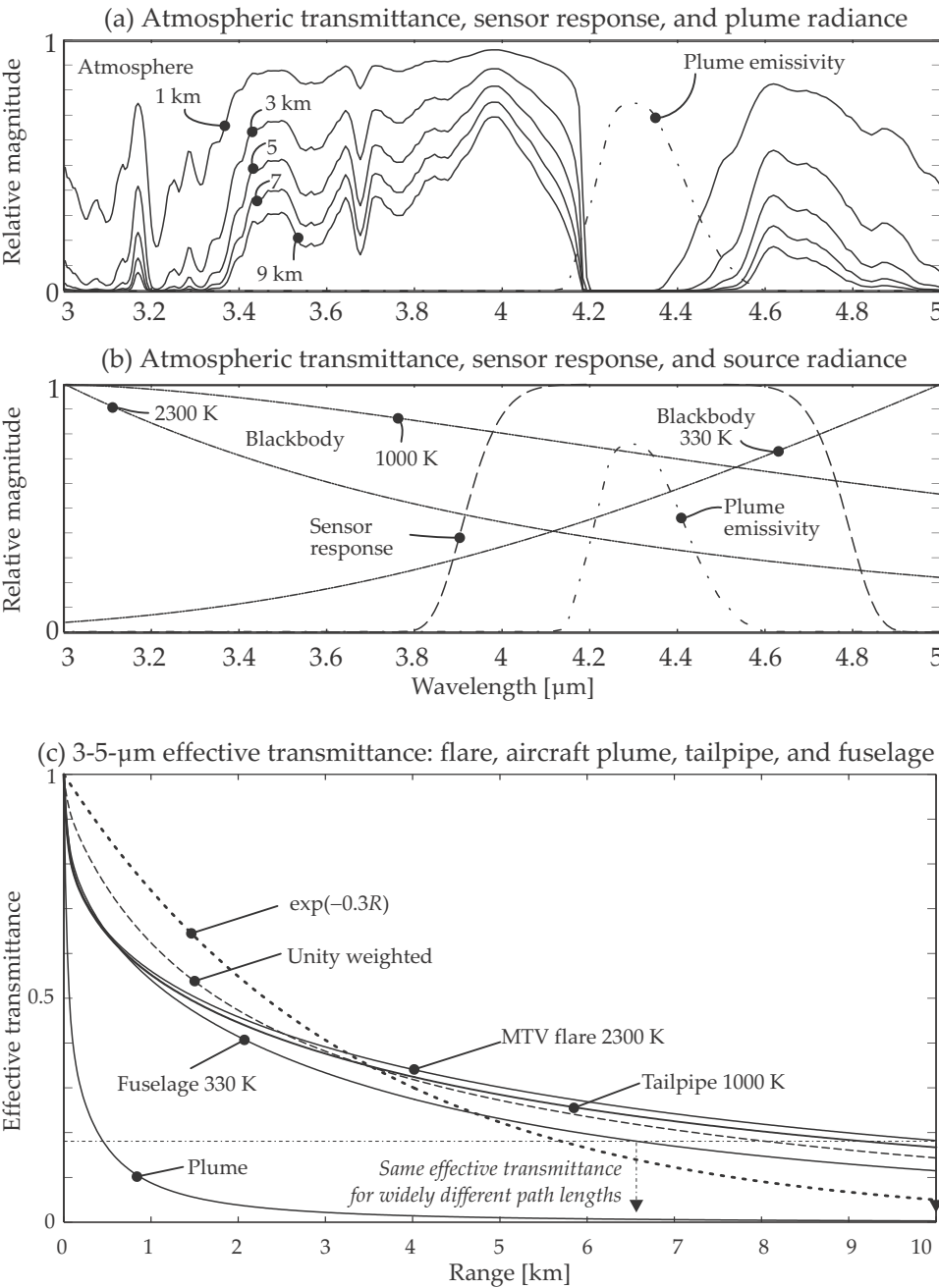


Figure 4.4 Effective atmospheric transmittance for various sources.

but only 6.5 km for a 330-K source. Hence, effective transmittance must be calculated for the correct source temperature.

It is often proposed that the effective transmittance curves versus distance can be approximated by an exponential curve $\tau_a = e^{-\alpha R}$. This formulation is Bouguer's law, as derived in Equation (4.4), at a single wavelength. When applied to broadband sensors, this approximation yields a poor fit for grey body radiation observed through a spectrally selective medium. Figure 4.4 also shows that Bouguer's law is totally inadequate in approximating the effective transmittance of a spectrally selective source through a spectrally selective medium.

4.5 Transmittance as Function of Range

It is sometimes necessary to scale transmittance from one path length to another transmittance at another path length. Equation (4.4) shows that for a homogeneous medium the spectral transmittance is given by

$$\tau(R) = e^{-\gamma R}, \quad (4.20)$$

where R is the path length in [m], and γ is the attenuation coefficient with units of $[\text{m}^{-1}]$. If the transmittance at range R_1 is known, the transmittance at range R_2 can be determined by realizing that the transmittance is characterized only by the attenuation coefficient

$$\gamma = -\log_e[\tau(R_1)]/R_1, \quad (4.21)$$

and hence

$$\begin{aligned} \tau(R_2) &= e^{-\gamma R_2} \\ &= e^{\log_e[\tau(R_1)]R_2/R_1}. \end{aligned} \quad (4.22)$$

The analysis in Section 4.4 indicates that scaling of effective transmittance with range strongly depends on the spectral radiance of the source. Hence, when scaling transmittance versus range, spectral transmittance should be used rather than wideband effective transmittance.

4.6 The Atmosphere as Medium

4.6.1 Atmospheric composition and attenuation

The atmosphere is a highly variable, complex, and dynamic medium.¹⁻³ The atmospheric density, pressure, and temperature vary with altitude, shown in Figure 4.5. The attenuation of electromagnetic radiation through the atmosphere is affected by absorption and scattering from molecules

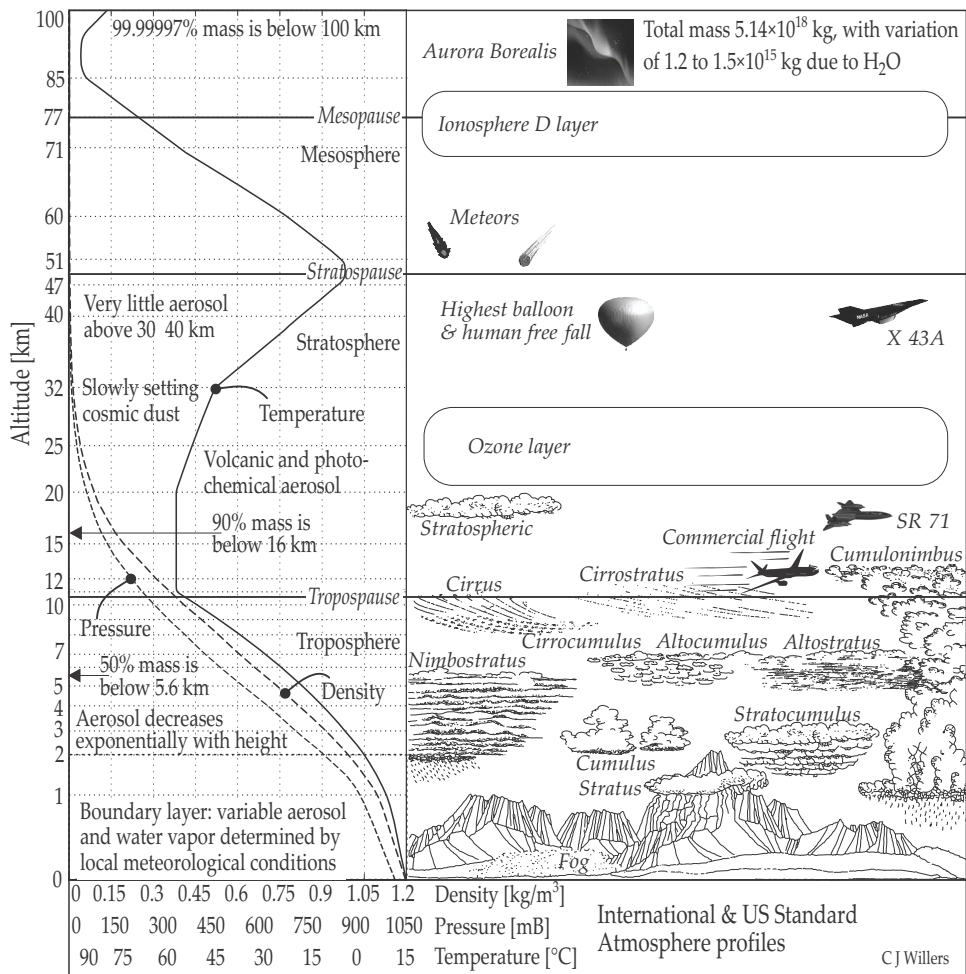


Figure 4.5 Atmospheric definitions and the standard atmospheric profiles.

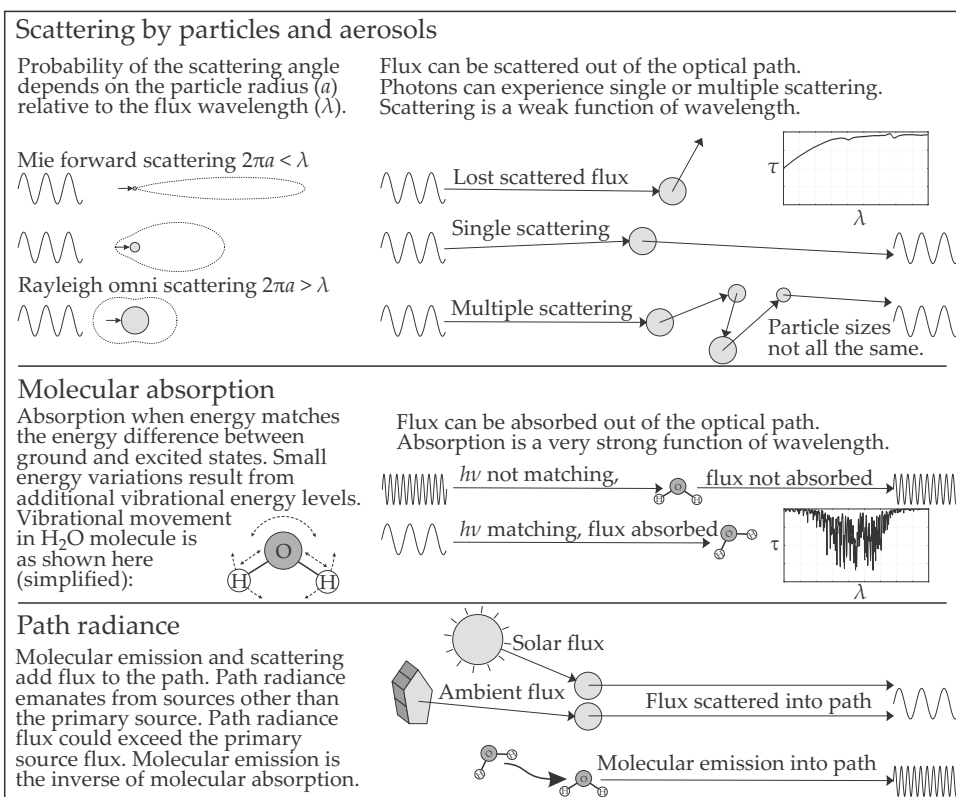


Figure 4.6 Atmospheric medium effects.

and particles (such as haze, dust, fog, or cloud droplets) suspended in the air. Figure 4.6 summarizes the medium effects of the atmosphere.

Scattering and absorption by aerosol particles is a prominent factor in the lower few kilometers in altitude, near the earth's surface. The molecular constituents in the atmosphere strongly affect attenuation and scattering in the visual spectral range. Atmospheric aerosol particles in the atmosphere vary greatly in their concentration, size, and composition, and consequently in their effects on visual and infrared radiation. The aerosols and molecules are not uniformly distributed along the optical path. Air density, and hence particle and aerosol density, decreases with altitude and even varies along paths at constant altitude. Some species of molecules may vary in concentration, depending on local conditions and air-mass history.

In terms of the model developed in Section 4.2.1, the atmospheric attenuation coefficient γ comprises two components $\gamma = \alpha + \sigma$, where α is the absorption coefficient, and σ is the scattering coefficient, both with

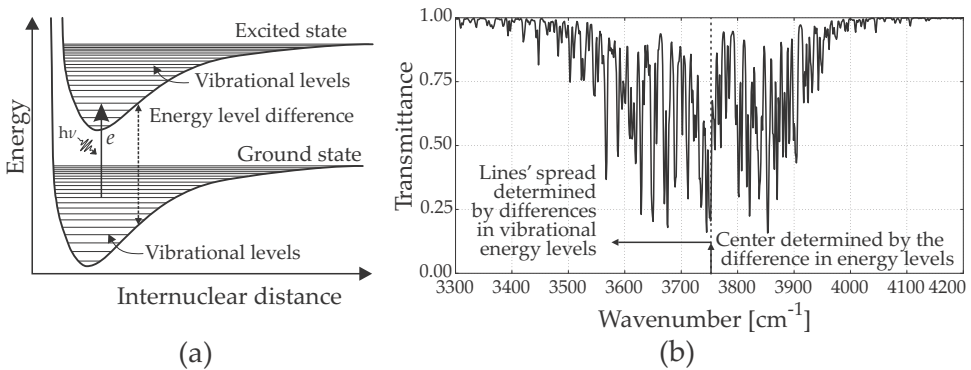


Figure 4.7 H_2O as a molecular absorber: (a) molecular energy levels and (b) atmospheric associated spectral transmittance over a 1-km path, near 3750 cm^{-1} .

units $[\text{m}^{-1}]$. α is a function of the molecular composition and density of the atmosphere, and varies strongly with wavelength. σ is a function of the aerosol and particulate contents of the atmosphere, and varies less strongly with wavelength.

4.6.2 Atmospheric molecular absorption

The atmosphere's main molecular constituents include molecular nitrogen (N_2 , 78%), molecular oxygen (O_2 , 21%), argon (Ar, 0.93%), molecular carbon dioxide (CO_2 , 0.04%), and water vapor (H_2O). Water vapor constitutes approximately 0.4% of the 'total' atmosphere but can range from 1% to 4% near the earth's surface. Molecular absorption in the infrared spectral band is however dominated by species with very low concentrations but very active vibration-rotation bands, such as water vapor, carbon dioxide, ozone, and nitrous oxide.²²

When a molecule absorbs a photon, an electron is excited from a lower energy level to a higher energy level. The energy levels are discrete quantum levels, but each quantum level has small variations due to molecule vibration, as shown in Figure 4.7(a). The diatomic molecules (N_2 , O_2 , CO) have one vibrational mode, known as vibrational stretch. These molecules lack a permanent dipole moment and are unable to sustain oscillating momentum, with less impact on transmittance and radiation. The triatomic molecules (CO_2 , N_2O , H_2O , O_3) have multiple vibrational modes, corresponding to low-energy level variations, within each quantum energy level. When a photon interacts with a molecule, the photon will be absorbed if the photon energy (hc/λ) matches one of the molecule's combined quantum and vibrational energy levels. Different molecules are in different vibrational energy states and hence absorb slightly different

wavelength photons, as shown in Figure 4.7(b).

Figure 4.8 shows the spectral transmittance of the atmosphere, but individually for different molecular species. Note how the different groups of lines correspond to different molecular quantum energy gaps. Within each group there are a multitude of minutely narrow lines due to vibrational differences between molecules.

4.6.3 Atmospheric aerosols and scattering

Aerosols in the boundary layer, up to 1–2 km in altitude, have greater variability compared to higher altitudes. These aerosols consist of a variety of natural and synthetic chemical compounds. Aerosol can be dry matter, liquid droplets, or a mixture of the two. Particle size can range from a few hundred nanometers to tens of micrometers. Aerosols are created by natural causes such as wind, volcanos, fog formation, fires, and even human and animal movement. Land aerosols comprise dust and organic particles from vegetation. Aerosols can also form by photochemical processes under solar irradiation. Heavier dust aerosols created mechanically, such as explosions, wind storms, or vehicle movement, typically contain a wide range of particle sizes. The composition of the dust aerosols depends on the means of activation, the properties of the soil, and the water content of the soil. The maritime aerosols are primarily salt particles and water spray. Anthropogenic causes include smoke and pollution aerosols from urban and industrial areas. Manmade aerosols have a different chemical composition compared to natural aerosols. Aerosols are mostly created near the earth's surface and are carried into the air by thermal and wind movement. Heavy aerosol particles (larger than a few μm) tend to drift to earth over time, whereas the lighter aerosols stay suspended in the air indefinitely or until settled down by rain. A heavy particle with diameter of 8 μm settles by gravitation²³ at an average velocity of 2 mm/s, but a 0.8- μm particle settles at a velocity of 0.02 mm/s.

Above the boundary layer, in the troposphere, the aerosols are less dependent on the local conditions near the earth's surface. At these altitudes, the different kinds of aerosol include smaller particles formed from gaseous components, undergoing processes of coagulation and agglomeration. There are also larger particles originating in the boundary layer, swept up by air currents. In the stratosphere (10–30 km), the aerosol distribution is globally more or less uniform. At these altitudes the aerosols are mostly sulfate particles formed by photochemical reactions. Volcanic eruptions can inject large volumes of dust and SO_2 aerosol, which are transported globally over very wide areas by stratospheric circulation. The

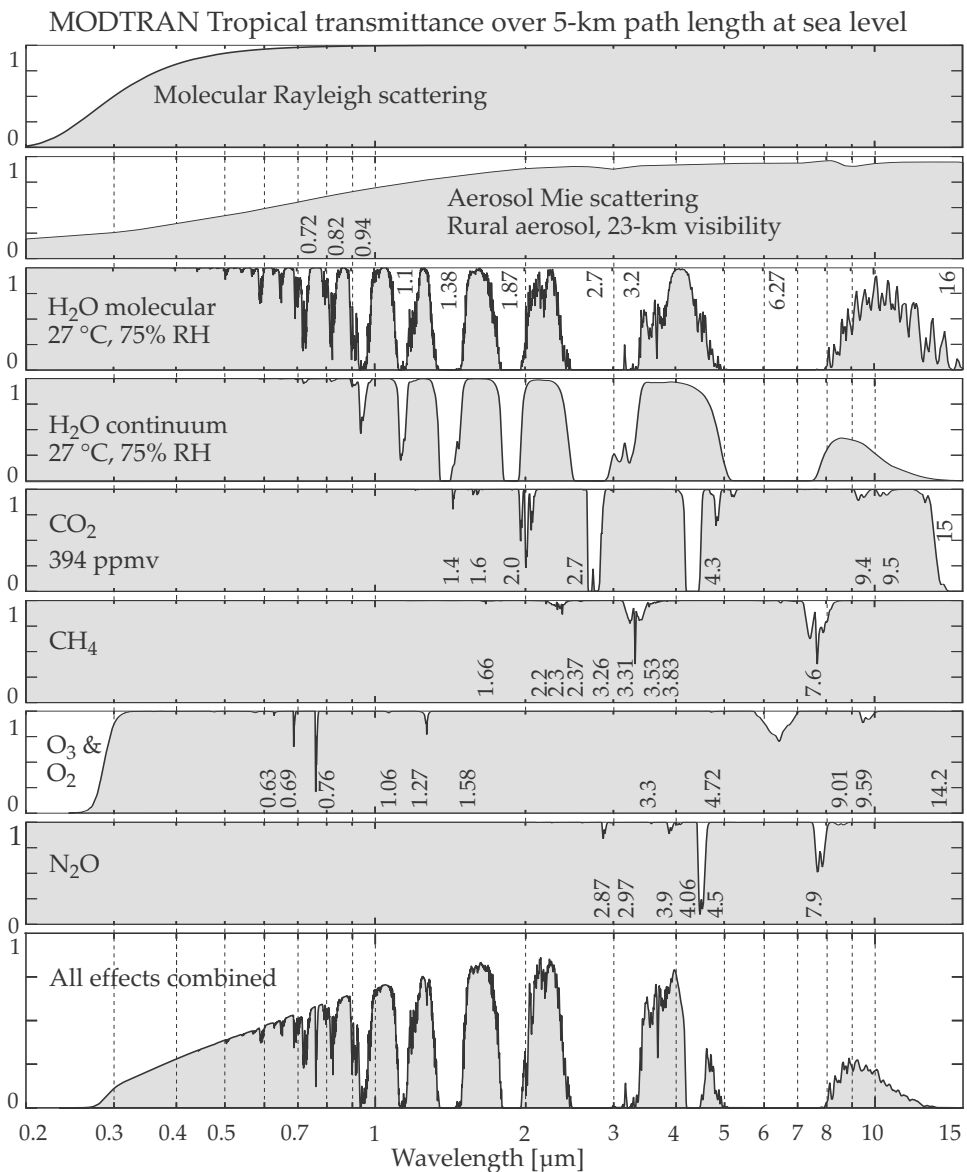


Figure 4.8 Transmittance of various aerosols and molecules in the atmosphere.

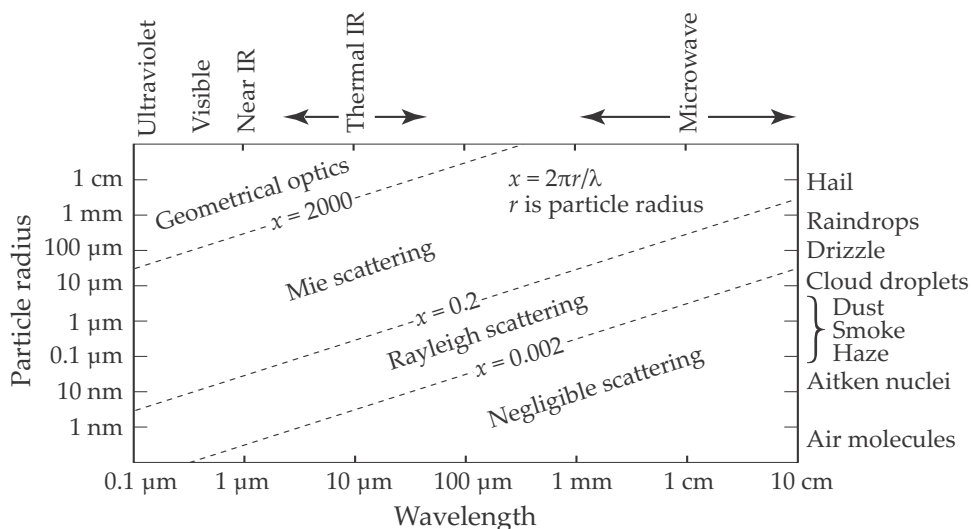


Figure 4.9 Aerosol scattering modes versus wavelength and particle size (used with permission¹).

heavier dust particles may drift down and settle out of the atmosphere in a few months, but photochemically formed sulphuric acid may remain for up to two years. Only a very small portion of the total aerosol content of the atmosphere exists above 30 km.²²

Aerosol composition is characterized with a statistical particle density versus a lognormal size distribution. The statistical modal value of radius is an indication of the aerosol 'size;' note, however, that the size has a statistical distribution with the modal size occurring most often, but there is a wide variation in particle sizes present. Indeed, most aerosols are characterized by several modal distributions, not just a single distribution. Hence, the concept of aerosol particle size should be understood to mean the most commonly occurring size in the distribution.

Scattering is a wideband phenomenon, compared to the narrow absorption lines of molecular absorption. The energy is momentarily absorbed and immediately re-radiated as if from a point source. The atmospheric effects on optical and infrared flux depend on the size of the particle compared to the wavelength of the light. Simplifying approximations are made to model the scattering effect. If $2\pi r/\lambda \ll 1$, where r is the particle radius, the Rayleigh approximation applies, and if $2\pi r \geq \lambda$, the Mie (sometimes called the Lorentz–Mie) approximation applies. Figure 4.9 shows the aerosol scattering modes at various wavelengths and particle sizes. Mie scattering is an approximation of scattering theory, and Rayleigh is a further special case of Mie scattering. Figure 4.10 shows scat-

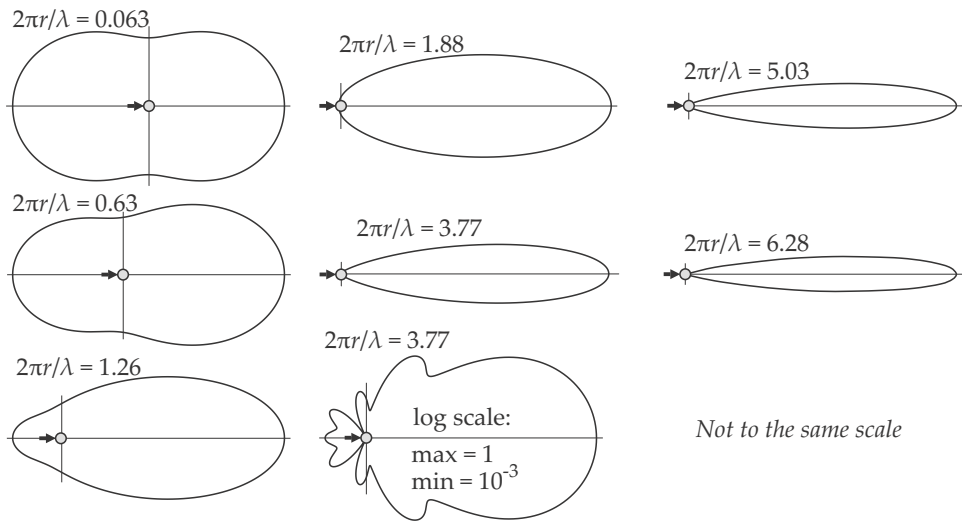


Figure 4.10 Scattering intensity profiles for various values of $2\pi r/\lambda$ in unpolarized sunlight.

tering intensity profiles for various values of $2\pi r/\lambda$ for a water droplet, starting with Rayleigh scattering and progressing to Mie scattering. Scattering also depends on the incident light polarization and on the geometrical shape of the particle. The discussion given here pertains only to unpolarized light and near-spherical and isotropic particles. Comprehensive coverage is available in Liou's book.⁴

Rayleigh scattering occurs when electromagnetic energy interacts with aerosol (molecules) of physical size significantly smaller than the wavelength of the energy field. This process is not a molecular absorption process but rather an interaction at energy levels other than molecular absorption. The resultant scattering attenuation coefficient γ is approximately proportional to λ^{-4} , affecting mainly propagation at shorter wavelengths (ultraviolet, visual, and near-infrared). Rayleigh scattering leads to the 'blue sky' observed visually. In the visual and ultraviolet bands, the blue-sky spectral radiance very roughly mimics a low-emissivity 10,000-K thermal radiator. Rayleigh scattering by atmospheric molecules has little effect at wavelengths longer than $3\ \mu\text{m}$. The phase function^{4,22} gives the probability distribution for the scattered light, so that $P(\theta)d\Omega$ is the fraction of the scattered radiation that enters a solid angle $d\Omega$ about the scattering angle θ . The phase function for Rayleigh scattering (for unpolarized light at visual wavelengths) can be approximated by $P(\theta) = k(1 + \cos^2 \theta)$, which has a component of omnidirectional scatter, with peaks in the forward and backward directions. Figure 4.8 shows Rayleigh molecular scattering over a 5-km path length.

Mie scattering occurs when electromagnetic energy interacts with particles of the same size or larger than the wavelength of the energy field, $2\pi r/\lambda \geq 1$, where r is the radius of the particle. *Mie* scatter has a large forward scatter along the direction of the incident flux. The shape of the phase function depends significantly on the physical and chemical characteristics of the aerosol. The resultant attenuation coefficient γ has some spectral variation but not as strong as Rayleigh scattering. Depending on the size of the particles and the flux wavelength, *Mie* scattering may have significant effects at wavelengths up to or even beyond 10 μm . For most naturally occurring low-density aerosols and artificial aerosols, the size distribution is such (average diameter less than 1 μm) that there is significant scattering in the visual region with minimal scattering in the infrared.

In less-dense mediums, a photon may only experience a single scattering event (single scatter) along the optical path. At higher aerosol densities, a single photon may experience many scattering events (multiple scatter). Clouds illuminated by sunlight have an intense white color; white because all colors are scattered equally (unlike Rayleigh's λ^{-4} spectral scattering), and opaque because of multiple-scattering of the sunlight. Figure 4.8 shows the transmittance along a 5-km path, with a 23-km visibility in a MODTRANTM Urban aerosol (see Section 4.7.2 for a description of MODTRANTM). The aerosol has an attenuating effect deep into the infrared spectral range.

4.6.4 Atmospheric transmittance windows

The atmosphere is transparent in some spectral bands; these bands are known as atmospheric windows. The transmittance in the atmospheric windows can be low, whereas in the spectral regions between the windows the transmittance is zero, truly opaque. The atmospheric windows are frequently named with the spectral band acronyms defined in Section 3.1.6.

Toward the shorter wavelengths, atmospheric absorption bands reduce the target flux by attenuation but do not contribute path thermal radiance flux. In the longer-wavelength bands, the atmosphere's effect in the absorption bands is twofold: a reduction in target flux as well as a strong thermal-path radiance contribution. Figure 4.11 shows how the radiance of a 300-K source contributes to the path radiance in the absorption bands beyond 5 μm . Keep in mind that the emissivity is $(1 - \tau)$ in the absorption bands! The designers of MWIR and LWIR systems therefore endeavor to limit sensor sensitivity in the absorption bands.

The visual spectral band, defined by human vision, is largely unaffected by molecular absorption. Atmospheric aerosol content is a limiting

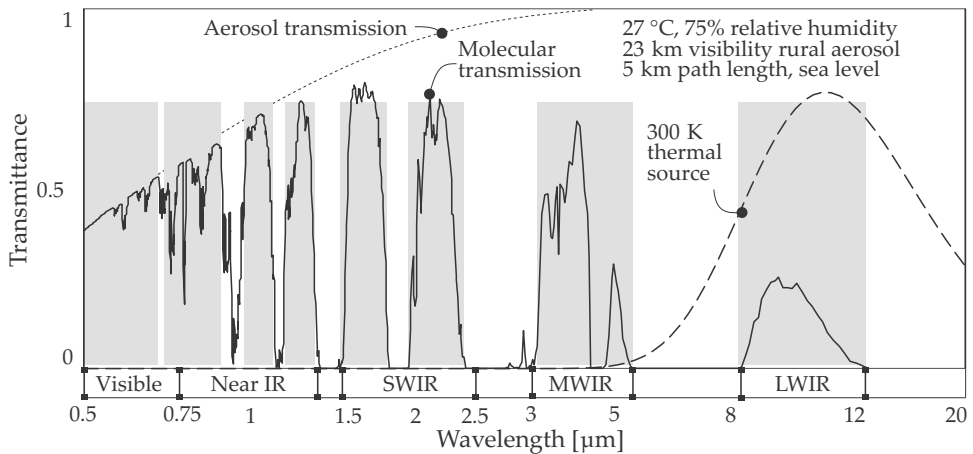


Figure 4.11 Atmospheric transmittance and atmospheric windows.

factor in this spectral band, ranging from Rayleigh scattering by molecules under clear-sky conditions to severe attenuation by heavy aerosol (cloud, dust) under poor visibility conditions.

The NIR spectral band has several narrow absorption spectral bands in the window. These absorption bands have no effect other than attenuating the target flux. Sensors operating in this band could use one or more of the narrow atmospheric windows.

The MWIR spectral band was traditionally used for the observation of hot targets (e.g., aircraft signatures). Recent detector developments provided staring array sensors with good performance against cooler ground targets as well. Sensor performance in this band is less sensitive to humidity, with the result that these sensors are used in humid/tropical areas. The sensors are not very sensitive in a low-ambient-temperature environment, with the result that these sensors are not very effective at the further northern and southern latitudes. The CO_2 absorption band at $4.3 \mu\text{m}$ severely attenuates flux but can at the same time be used to detect hot gas CO_2 emissions. Sensor design sometimes uses a narrower portion of the band and not the full width of the atmospheric window.

The LWIR spectral band is commonly used for the observation of ground-based targets. Ground targets have temperatures around 300 K, which result in peak infrared exitance and strong signals in this atmospheric window. Sensor performance in this band is sensitive to humidity, with the result that these sensors are mainly used under cooler and drier climatic conditions and less often in high humidity climates. Diffusely reflected sunlight from high-emissivity (low reflectivity) surfaces has no appreciable effect in the 8–12- μm spectral band; the observable signature

stems mainly from the object's thermal exitance. Sunlight reflection from specular surfaces (known as glint) may produce a significant signal in the LWIR band.²⁴ See also Sections 8.1 and 8.11 and Figure 8.3 for a discussion of the effect of sunlight and sunglint on optical signatures.

4.6.5 Atmospheric path radiance

Atmospheric path radiance has a major influence on sensor performance and should be considered alongside atmospheric transmittance in system design. Equation (4.10) provides a valuable insight into the concept of a medium's path radiance, but it remains a simplified model. The scope of its validity will now be reviewed. For a viewer on the ground, the zenith angle is defined as the angle between the direction of view and the vertical. The atmospheric transmittance [Figure 4.12(a)] and path radiance were calculated for a slant path to space for zenith angles ranging from the vertical to the horizontal. The MODTRANTM computer code was used to calculate the transmittance and path radiance for the Tropical atmosphere. This atmospheric model has a temperature of 300 K at ground level, decreasing with altitude. Figure 4.12(b) shows the MODTRANTM path radiance as well as the path radiance predicted by Equation (4.10). Four spectral regions were used: the 1.5–2.5 μm , 3–5 μm , and 8–12- μm spectral windows, as well as a 6–7- μm band in an absorption spectral region. At near zero zenith angles, the simple model overpredicts path radiance for the 3–5- μm and 8–12- μm spectral windows. For the 6–7- μm spectral band the fit is perfect, and for the 1.5–2.5- μm spectral band the prediction is not even on the graph. At near-horizontal views, the simple model predicts more accurately for the three thermal bands.

The simple model in Equation (4.10) requires two conditions for accuracy: a dense medium and a homogeneous medium. Section 4.2.5 shows that an optically thick gas medium has a high emissivity, and acts like a surface radiator. A horizontal atmospheric slant path to space is reasonably optically thick, and the temperature is constant for a considerable distance, hence the approximation is good. For zero zenith angles the path is initially optically thick, but with increasing altitude the atmospheric density decreases, as does the temperature, resulting in a somewhat poorer fit. At small zenith angles the observer is looking through a warmer atmosphere at low altitudes and at a colder atmosphere at higher altitudes, hence the lower path radiance. The simple model, as used here, only includes thermal radiation, and hence it is not suitable for modeling path radiance in the 1.5–2.5- μm spectral band.

For systems operating in the visual or NIR part of the spectrum the

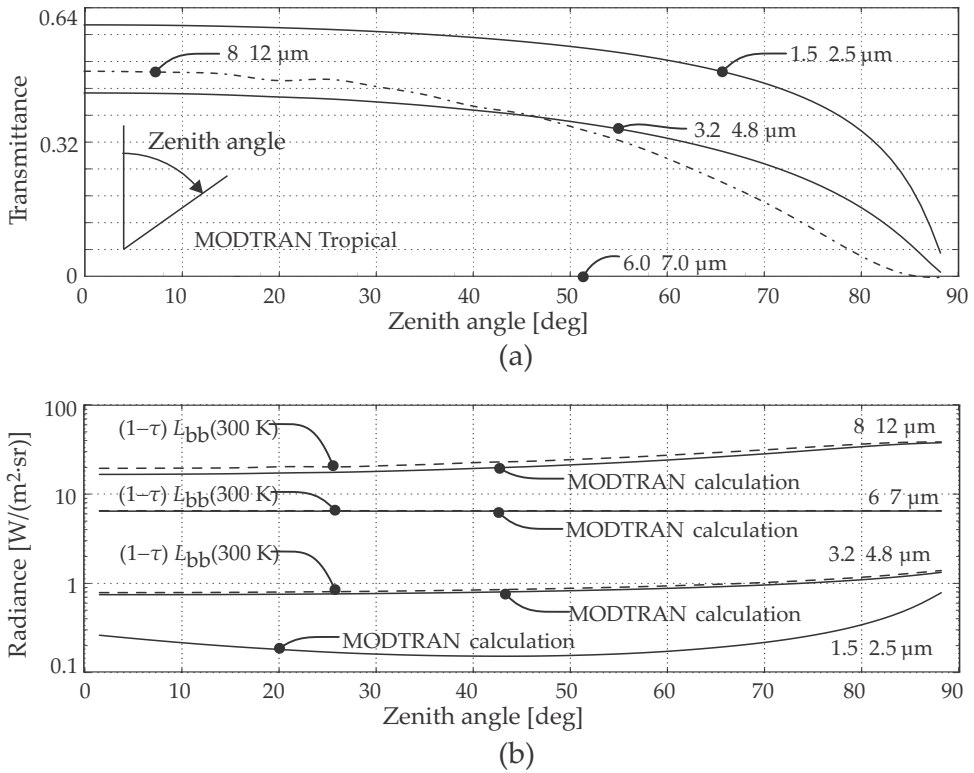


Figure 4.12 Atmospheric transmittance and path radiance to space, actual and simplified model: (a) transmittance vs. zenith angle and (b) path radiance vs. zenith angle.

self-emission term L_{th} will be zero, so that

$$L_R = L_0 e^{-\gamma \bar{R}} + \frac{\sigma L_\sigma}{\gamma} (1 - e^{-\gamma \bar{R}}). \quad (4.23)$$

Duntley¹⁷ reports that experiments confirmed that Equation (4.23) is valid for visual observation provided that the parameters are all weighted with the visual response, and that the radiance terms are replaced with luminance terms.

For systems operating in the MWIR spectral range all of the terms in Equation (4.10) must be kept because both particulate scattering and path emission are present, especially under low-visibility (less than 5 km) conditions:

$$L_R = L_0 e^{-\gamma \bar{R}} + \frac{\sigma L_\sigma + \epsilon L_{th}}{\gamma} (1 - e^{-\gamma \bar{R}}). \quad (4.24)$$

Due to all of the approximations made, Equation (4.24) is probably least accurate in this spectral region, and if accurate calculations must be performed, the differential equations should be solved numerically.

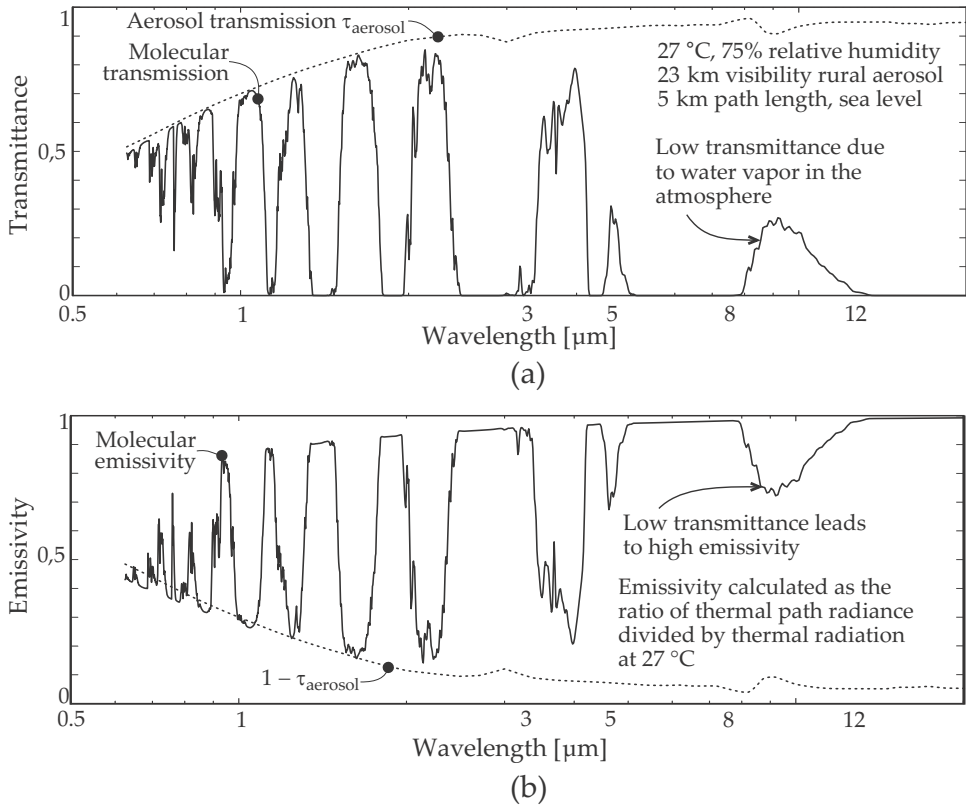


Figure 4.13 Atmospheric transmittance and emissivity: (a) MODTRAN™-calculated transmittance and (b) emissivity derived from MODTRAN™-calculated path radiance.

For systems operating in clear air in the LWIR spectral range the scattering coefficient σ and the term L_σ will be zero:

$$L_R = L_0 e^{-\gamma \bar{R}} + \frac{\epsilon L_{\text{th}}}{\alpha} (1 - e^{-\gamma \bar{R}}) \quad (4.25)$$

$$= L_0 e^{-\gamma \bar{R}} + L_{\text{th}} (1 - e^{-\gamma \bar{R}}). \quad (4.26)$$

For systems operating in severe aerosol both terms must be retained, and the same reservation applies as for MWIR systems.

4.6.6 Practical consequences of path radiance

Figure 4.13(a) shows the spectral transmittance for a MODTRAN™ Tropical atmosphere for a 5-km path length. The effect of aerosol scattering in the visual spectral band is clearly visible. The transmittance graph applies to a 5-km path at 23-km visibility — for poorer visibility, the transmission is even less than shown here. In the infrared spectrum, carbon dioxide

and water vapor are particularly relevant. Because these molecules absorb infrared energy, they also emit infrared energy.

The emissivity [Figure 4.13(b)] was determined as the ratio of the Tropical MODTRANTM predicted path radiance (thermal component only), and a thermal radiator at 300 K (27 °C):

$$\epsilon = \frac{L_{\text{path thermal}}}{L_e(300 \text{ K})} , \quad (4.27)$$

where the path radiance $L_{\text{path thermal}}$ was calculated by MODTRANTM, and L_e is Planck's law for a 300-K source (the temperature of the Tropical atmosphere). In the thermal spectral bands, this is a perfectly legal operation, as defined in Equation (3.24).

Figure 4.13 shows that, even for the Tropical atmosphere at 27 °C and 75% relative humidity, a 5-km path length results in an atmospheric emissivity of 0.75 in the 8–12- μm band. An 8–12- μm thermal imager is therefore trying to observe the target, looking through a veiling hot blackbody with an emissivity of 0.75. The water vapor has much less effect on transmittance in the 3–5- μm spectral band. The emissivity in the 3–5- μm band is affected by CO₂ at 4.3 μm .

4.6.7 Looking up at and looking down on the earth

Figure 4.14 shows the downward and upward radiance along a slant path. The MODTRANTM Tropical model and Rural aerosol, with 23-km visibility, was used in this calculation. The path radiance term includes thermal path radiance and single-scattered sunlight path radiance. The total path length is 11.3 km, with its two endpoints at sea level and 8 km. Note the atmospheric absorption around 4.3 μm . The MODTRANTM Tropical model has a sea-level temperature of 300 K. When looking down, the warm terrain is observed in regions with good transmittance, and the cold atmosphere is observed in spectral regions with poor transmittance. When looking up, the warm atmosphere is observed in spectral regions of poor transmittance (i.e., high emissivity), and the cold space is observed in spectral regions with good transmittance. The positive–negative relationship between the looking-up and looking-down curves illustrate the principle of atmospheric exitance in the absorption bands. These MODTRANTM predictions agree well with published measured data.¹

4.6.8 Atmospheric water-vapor content

Water vapor in the atmosphere is a very important consideration during infrared system design. Figure 4.15 shows the vertical water-vapor content

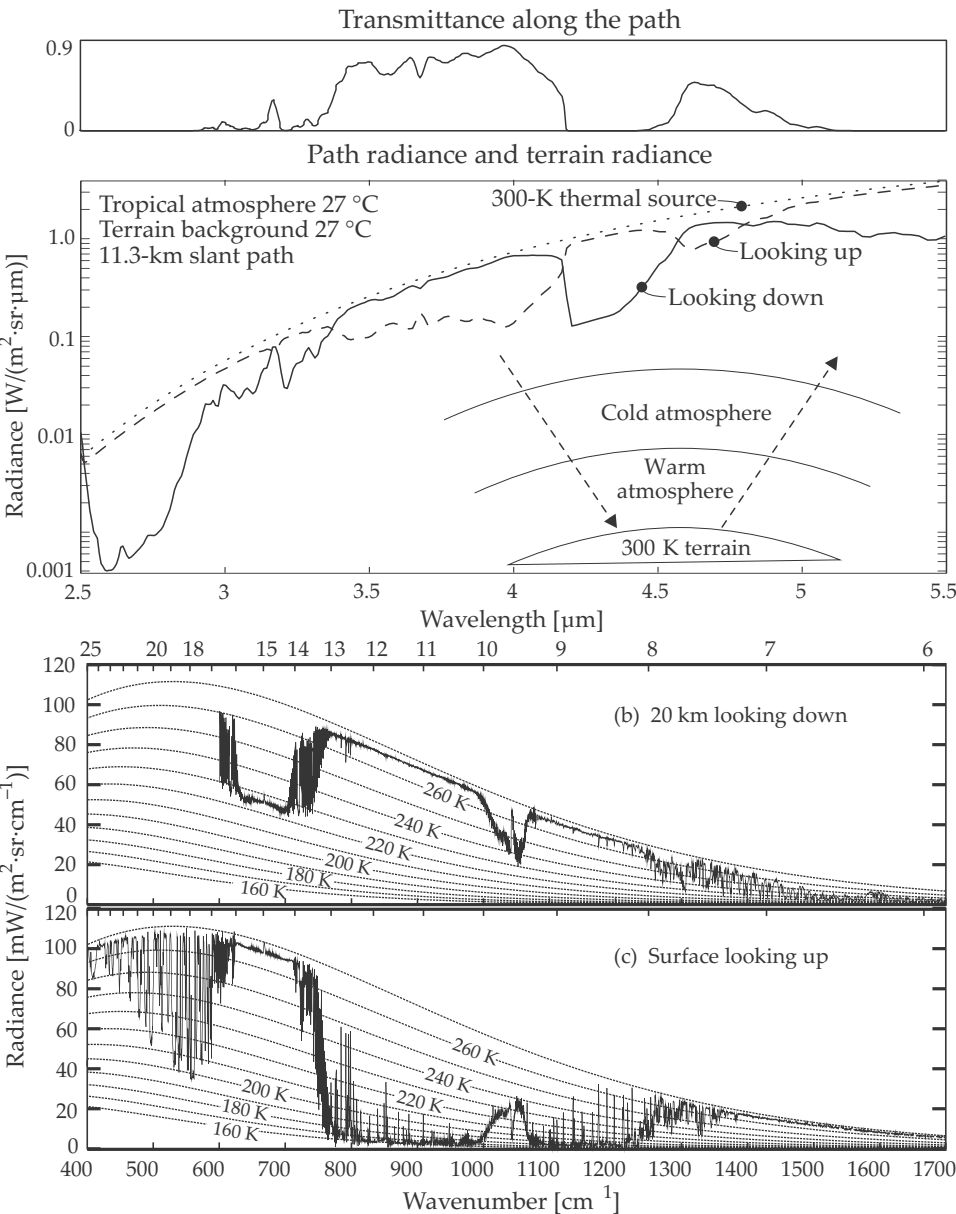


Figure 4.14 Radiance up and down along a slant path: (a) MODTRAN™ calculation, (b) measured (looking down), and (c) measured (looking up). Measured graphs (b), and (c) used with permission.¹

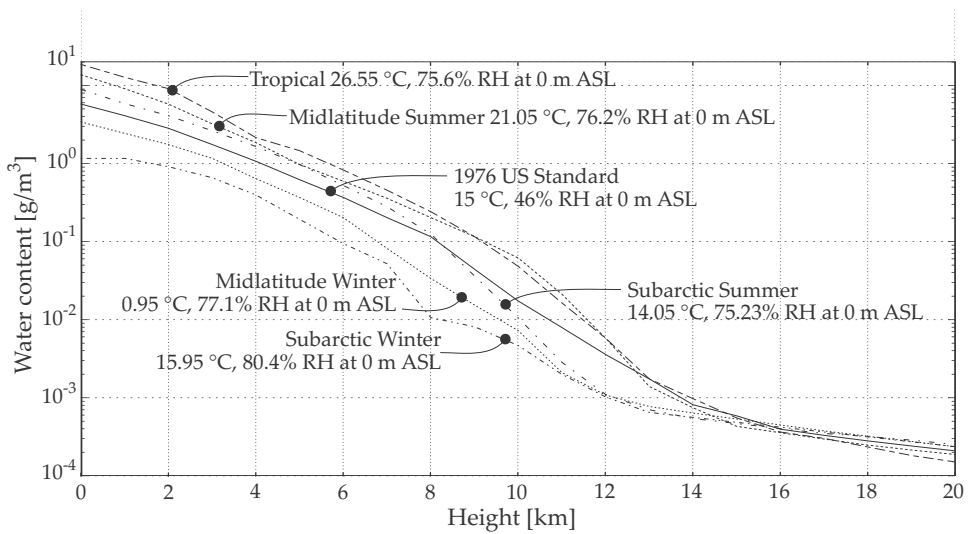


Figure 4.15 MODTRAN™ atmospheric water vapor content profiles.

for the standard MODTRAN™ atmospheric models. These profiles represent ‘typical conditions’ rather than the extremes that may occur. Also shown in the figure are the temperatures and relative humidity values for the models at sea level. Note the wide variability in water vapor in the troposphere but the relatively homogenous content in the stratosphere.

Atmospheric humidity is commonly expressed in relative humidity (RH). Relative humidity is the quantity of water vapor in the atmosphere, expressed as a percentage of the maximum absolute humidity. The maximum absolute humidity is determined by the water-vapor partial pressure at the atmospheric temperature and is given by

$$q = \frac{1325.252}{T} 10^{7.5892(T-273.15)/(T-32.44)}, \quad (4.28)$$

in units of $[\text{g}/\text{m}^3]$ and temperature in $[\text{K}]$. This equation has an error of less than 1% over the range $-20\text{ }^{\circ}\text{C}$ to $0\text{ }^{\circ}\text{C}$, and less than 0.1% for the temperature range $0\text{ }^{\circ}\text{C}$ to $50\text{ }^{\circ}\text{C}$.

Figure 4.16 shows the atmospheric absolute humidity versus temperature. Note how sharply the absolute humidity increases with increasing temperature. At $30\text{ }^{\circ}\text{C}$ the atmospheric water content is almost double the atmospheric water content at $20\text{ }^{\circ}\text{C}$. Figure 4.16 also shows the effective transmittance for a 5-km atmospheric path with relative humidities of 60% and 95%, at various temperatures. For example, an atmosphere with 60% relative humidity at a temperature of $30\text{ }^{\circ}\text{C}$ has an effective transmittance of 0.21. The highest recorded²⁵ absolute humidity was $37.5\text{ g}/\text{m}^3$ in Sharjah, United Arab Emirates.

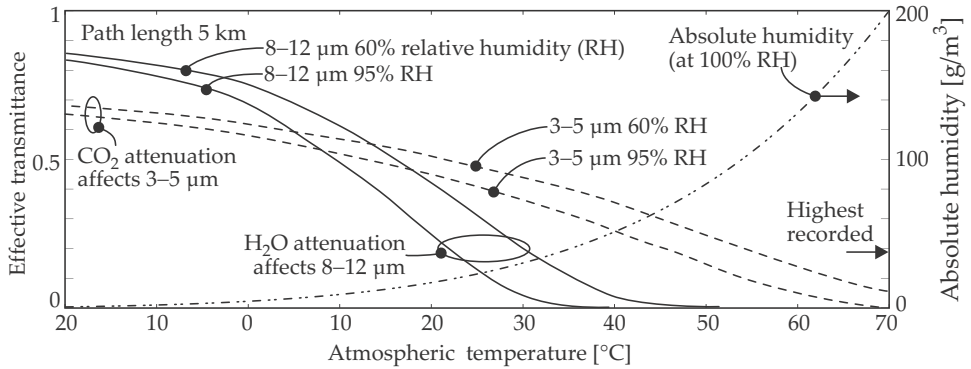


Figure 4.16 Effective atmospheric transmittance for a humid atmosphere: transmittance on left, and absolute humidity on right.

4.6.9 Contrast transmittance in the atmosphere

Section 4.2.3 derives a general law for contrast reduction. This section investigates contrast transmittance for paths in the atmosphere. Three path types are considered here: observations along an upward path, a downward path, and a horizontal path.

4.6.9.1 Upward observations

In Equation (4.16) the E_{b0} term is the sky radiance along the line of sight as measured from the observer (to space), and E_{bR} is the sky radiance along the line of sight as measured from the object (to space). Combining Equations (4.10) and (4.16), the contrast transmittance for upward observations is obtained as

$$\tau_c = \frac{C_R}{C_0} = e^{-\gamma \bar{R}} \left[\frac{1 - e^{-\gamma \bar{R}_{R\infty}}}{1 - e^{-\gamma \bar{R}_{0\infty}}} \right], \quad (4.29)$$

where $\bar{R}_{R\infty}$ denotes the integral $\bar{R} = \int_R^\infty f(r)dr$ along the line of sight, and $\bar{R}_{0\infty}$ is the integral $\bar{R} = \int_0^\infty f(r)dr$ [see Equation (4.6)]. For an upward path the term in the square brackets is almost always less than unity because there is a shorter path in the numerator and a longer path in the denominator. If the atmosphere along the path is uniform (which the atmosphere is not), these integrals degenerate to the path lengths because $f(r)$ is the same constant value in both cases. The integration limit ∞ is not really infinity because the atmosphere has finite extent. The symbol ∞ in this context only indicates that the path runs into space.

Two important observations can be made from Equation (4.29): (1) τ_c depends only on γ and \bar{R} . In an atmosphere isotropic with respect

to azimuth angle, \bar{R} only depends on the elevation angle, i.e., the angle rising above the horizon. Consequently, for an object moving at a constant elevation angle, the atmosphere will always have a constant contrast transmittance. (2) The contrast transmittance is independent of the path radiance phenomenon, scattering, or thermal emission.

4.6.9.2 Downward observations

Equations (4.10) and (4.16) can be combined¹⁷ to yield

$$\tau_c = \frac{1}{1 - \frac{\sigma L_\sigma + \epsilon L_{\text{thermal}}}{\gamma L_0} (1 - e^{\gamma \bar{R}})}, \quad (4.30)$$

where the variables are defined in Section 4.2.2. Note that all of the variables are properties of the medium except L_0 , which is a property of the source. This equation applies for both scattered and thermally radiated path radiance components.

In the *visual spectral band* the thermal radiance of the atmosphere is insignificant, and Equation (4.30) can be simplified to

$$\tau_c = \frac{1}{1 - \frac{\tau L_\sigma}{\gamma L_0} (1 - e^{\bar{R} \gamma})} \quad (4.31)$$

$$= \frac{1}{1 - K_\nu (1 - e^{\bar{R} \gamma})}, \quad (4.32)$$

where K_ν is called the *sky-ground ratio* and to some degree resembles the ratio of radiance values of the sky (as seen from the ground) and the ground radiance (as seen from the sky).²⁶ Typical values of K_ν are given in Table 4.1. Note that these values are only applicable to observations in the visual spectral band. Duntley¹⁷ and Gordon¹⁸ describe how to determine K_ν . The sky-ground ratio has a very strong influence on the contrast transmittance τ_c .

In the *MWIR spectral band*, Equation (4.30) cannot be simplified because the scattered and thermal path radiance components are similar in magnitude.

In the *LWIR spectral band*, the scattering of the clear sky atmosphere is insignificant, and Equation (4.30) can be likewise simplified. It is convenient to define a sky-ground ratio for the infrared domain,²⁶

$$K_\mu = \frac{\epsilon L_{\text{th}}}{\gamma L_0}. \quad (4.33)$$

Typical values of the infrared sky-ground ratio are shown in Table 4.2. If the source object is near the ground, L_0 is the earth radiance, and the

Table 4.1 Sky–ground ratios in the visual spectral band.

Sky condition	Ground condition	K_v
overcast	fresh snow	1
overcast	desert	7
overcast	forest	25
clear	fresh snow	0.2
clear	desert	1.4
clear	forest	5

Table 4.2 Sky–ground ratios in the infrared spectral bands.

Spectral band	K_μ	K_μ
3–5 μm	0.70	1.42
8–14 μm	0.85	1.17
10–12 μm	0.86	1.17
	–10 K	+10 K

downwards atmospheric path is short. If the atmospheric temperature is similar to the earth’s temperature, the path radiance in opaque spectral bands is similar to the terrain radiance. Hence, the observer sees the same radiance, be it terrain or atmosphere.

The upward path to space through the atmosphere has relatively low transmittance, with most of the lower atmosphere contributing to the path radiance.²⁷ K_μ can therefore be approximated by the ratio of the atmospheric boundary layer path radiance to the earth radiance. For look-down observations, at targets near the earth’s surface, the atmosphere’s contrast transmittance is then

$$\tau_c = \frac{1}{1 - \frac{L_o(T_{air})}{L_o(T_{earth})}(1 - e^{\bar{R}\gamma})}. \tag{4.34}$$

The background radiance relative to the atmospheric radiance depends on the earth and atmospheric temperatures. In the equilibrium situation, when the earth temperature equals the air temperature, the ratio of exitance values is unity, and thus K_μ is unity. This is however a much-simplified case because there is always a difference between the temperatures of objects and the air. The differences depend on the time of day, wind velocity, material properties of the object, etc. Suits²⁸ estimates that typical maximum differentials between the air and earthbound objects are in the region of 10 K. Little K_μ data is published in the literature. For hy-

pothetical systems with square spectral passbands, the approximate values for K_μ are listed in Table 4.2.

4.6.9.3 Horizontal observations

For horizontal paths the contrast transmittance is the limiting case of both Equations (4.29) and (4.30), with the same result. For shorter distances with small R , $e^{-\gamma \bar{R}_{R\infty}} \approx e^{-\gamma \bar{R}_{0\infty}}$ in Equation (4.29). For longer distances $e^{-\gamma \bar{R}} \rightarrow 0$. In Equation (4.30) the background exitance is replaced by the sky exitance, and K_V and K_μ become unity. Therefore, on the horizon, the contrast transmittance is equal to the radiance transmittance:

$$\tau_c = e^{-\bar{R}\gamma}. \quad (4.35)$$

4.6.10 Meteorological range and aerosol scattering

Aerosol types are often characterized by a range parameter called ‘meteorological range’ or ‘visibility’. This parameter is defined as the range where the atmosphere reduces the apparent contrast of a unity contrast target to the foveal contrast threshold (C_v) of the human eye. In other words, how far can you see a large black object against a white background? A reference wavelength of 550 nm is used. In 1924, Koschmieder determined C_v to be 0.02. From Equation (4.35) it follows that the meteorological range R_V is given by

$$R_V = \frac{-\ln(0.02)}{\sigma_{550 \text{ nm}}} = \frac{3.91}{\sigma_{550 \text{ nm}}}, \quad (4.36)$$

where $\gamma = \sigma_{550 \text{ nm}}$ is the aerosol scattering coefficient. Note that R_V and $\sigma_{550 \text{ nm}}$ must be in the same units.

The World Meteorological Organization determined a different contrast threshold,²⁹ $C_V = 0.05$; with the new value being empirically determined for human observations in the real world. The value $C_v = 0.05$ leads to the form $R_V = 3/\sigma$. Note that both forms are referred to by the same name but yield different meteorological ranges. The Koschmieder range is 30% further! It is therefore important to state which convention is used. The MODTRANTM aerosol-model meteorological range is defined in the original Koschmieder $C_v = 0.02$ convention.³⁰ Hence, the value to be used in MODTRANTM must be 1.3 ± 0.3 times the real-world human-eye observation distance. See Section 9.2 for a numerical calculation of meteorological range using a simple radiometric model.

Figure 4.17 shows the spectral scattering coefficient for various MODTRANTM aerosols. Also shown are the approximate aerosol modal diam-

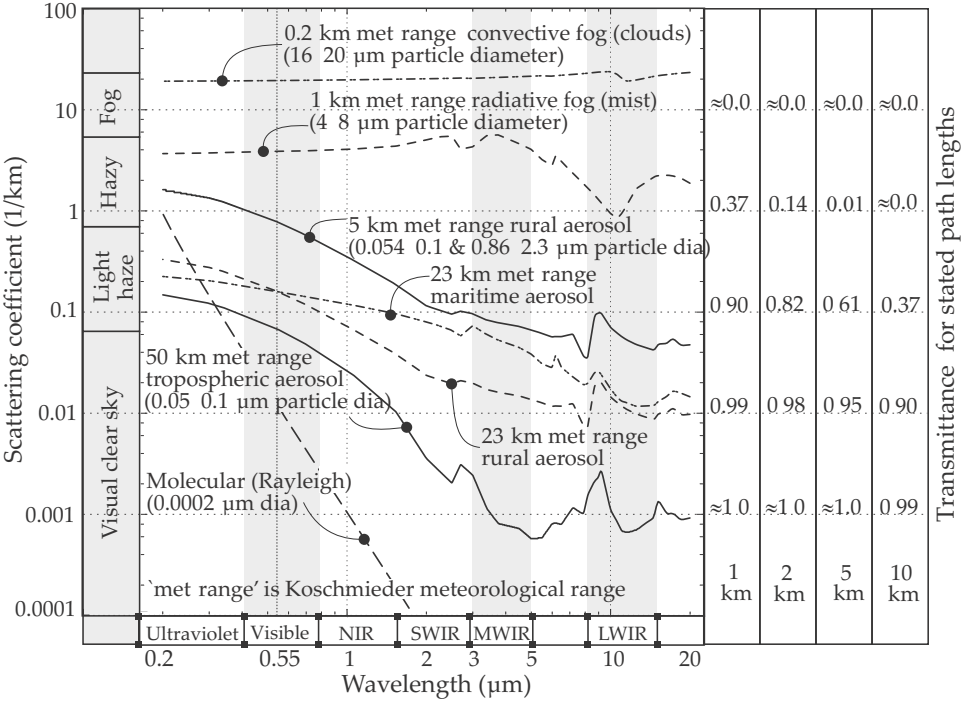


Figure 4.17 Spectral scattering attenuation coefficient for MODTRAN™ aerosols and indicative transmittance for various path lengths.

eters (most-common particle size for the specific aerosol). Shown on the right side of the graph is the transmittance over a 1-km path, corresponding to the scattering coefficient on the left side.

In Figure 4.17, from the bottom to the top, the aerosol size varies from small to large. Clear-sky conditions (blue-sky Rayleigh scattering and the 50-km meteorological range tropospheric atmosphere) affects the ultraviolet, visible, and near infrared spectral bands. Light haze conditions (23-km meteorological range) affect the MWIR and LWIR bands to some extent. Finally, the large particle fog aerosol affects all of the visual and infrared bands equally.

Meteorological range is defined in terms of human vision, not infrared terms. Two aerosols may have the same meteorological range, but have very different infrared scattering properties. Compare the 23-km Rural and Maritime MWIR and LWIR aerosol scattering coefficients in Figure 4.17. Even though the scattering coefficient is the same at 550 nm, it differs by a factor of five in the MWIR band.

4.7 Atmospheric Radiative Transfer Codes

4.7.1 Overview

Atmospheric radiative transfer codes³¹ are computer programs that calculate properties of the atmosphere, pertaining to radiative flux transfer. The codes are typically designed and focused toward a specific application, but some codes have more general applicability. Some codes provide specialist additional functionality such as solar or lunar irradiance or background scene images. The codes are mostly discrete ordinate models (see Section 4.3). In this section the MODTRANTM code is briefly summarized.

4.7.2 MODTRANTM

MODTRANTM (MODerate resolution TRANsmission) is an atmospheric code to calculate the direct and diffuse transmission, the path radiance, transmitted and top-of-atmosphere solar/lunar irradiances, and more, for a specified path through the atmosphere. MODTRANTM can be used as a stand-alone program, but it can also be interfaced as a subroutine to larger software systems. MODTRANTM is available as source code or a compiled binary file.³² From the MODTRANTM website:²¹

MODTRANTM is a 'narrow band model' atmospheric radiative transfer code. The atmosphere is modeled as stratified (horizontally homogeneous), and its constituent profiles, both molecular and particulate, may be defined either using built-in models or by user-specified vertical profiles. The spectral range extends from the UV into the far-infrared ($0\text{--}50,000\text{ cm}^{-1}$), providing resolution as fine as 0.2 cm^{-1} . MODTRANTM solves the radiative transfer equation including the effects of molecular and particulate absorption/emission and scattering, surface reflections, and emission, solar/lunar illumination, and spherical refraction.

The atmosphere is modeled as 36 discrete layers from sea level up to 100-km altitude. Layer thickness is not constant — it is smaller in the lower atmosphere and increases in the upper atmosphere. Each layer is modeled as a homogenous medium with an appropriate temperature, pressure, molecular composition, and aerosol distribution. Six standard atmospheric models are provided, with the vertical profiles shown in Figure 4.18. The figure also shows the altitudes of the MODTRANTM layers. The user can add detailed new atmospheric models. These models need not comply with the MODTRANTM discrete layer definitions. The user can also specify

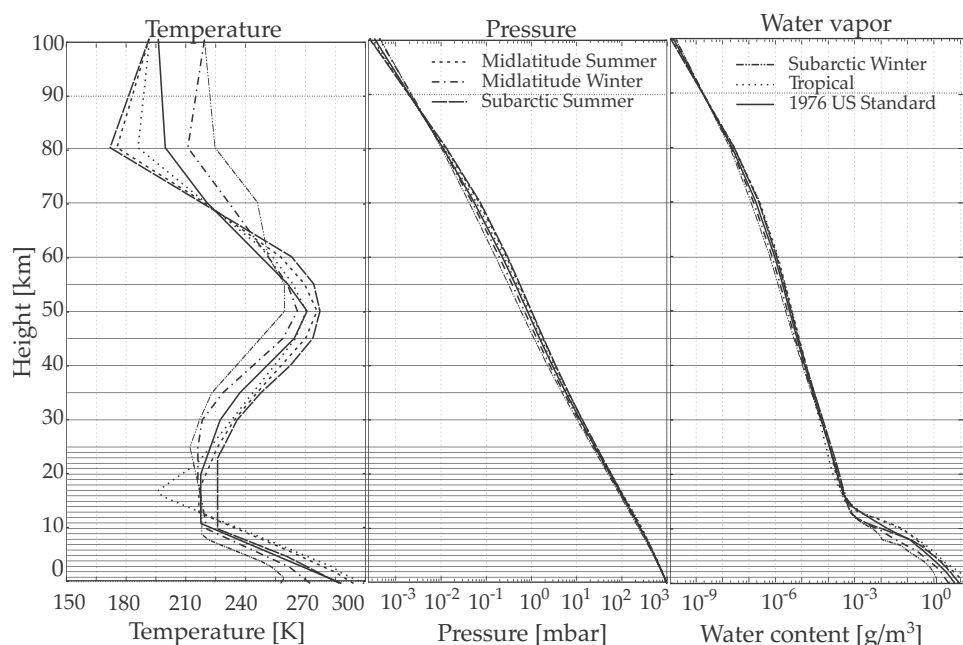


Figure 4.18 Standard MODTRAN™ atmospheric models' discrete layer altitudes and vertical profiles.

custom aerosol distributions.

Transmittance is calculated using molecular absorption coefficients, line density parameters, and average absorption line widths, all of which are temperature dependent over atmospheric temperature ranges. Aerosol scattering is calculated using aerosol size distribution and humidity. Standard MODTRAN™ aerosol models include Urban, Rural, Maritime, Fog, and Rain, which also provide for seasonal variation in the aerosol distribution at higher altitudes. Molecular continuum absorption, molecular scattering, and aerosol absorption and scattering are also calculated. MODTRAN™ calculates the path radiance component from thermal self-emission, solar and lunar scatter into the path, direct solar irradiance, and multiple-scattered solar or self-emitted radiance.

Path geometry calculation includes the earth curvature spherical refraction, accurately calculated as a function of zenith angle. The atmospheric amounts (molecular and aerosol) are calculated along the slanted path lengths in each layer of the model. The code makes provision for horizontal paths, slanted paths, paths to space, and paths from space. The user has several alternative options to define the paths.

Bibliography

- [1] Petty, G. W., *A First Course in Atmospheric Radiation*, Sundog, Madison, WI (2006).
- [2] Farmer, W. M., *The Atmospheric Filter: Volume I Sources*, JCD Publishing, Winter Park, FL (2001).
- [3] Farmer, W. M., *The Atmospheric Filter: Volume II Effects*, JCD Publishing, Winter Park, FL (2001).
- [4] Liou, K. N., *An Introducton to Atmospheric Radiation*, Academic Press, San Diego, CA (2002).
- [5] Kondratyef, K. Y., Ivlev, L. S., Krapvin, V. F., and Varatsos, C. A., *Atmopsheric Aerosol Properties*, Springer Praxis, Berlin (2006).
- [6] Bohren, C. F. and Clothiaux, E. E., *Fundamentals of Atmospheric Radiation: An Introduction with 400 Problems*, Wiley-VCH, New York (2006).
- [7] Wyngaard, J. C., *Turbulence in the Atmosphere*, Cambridge University Press, Cambridge, UK (2010).
- [8] Smith, F. G., Ed., *The Infrared and Electro-Optical Systems Handbook: Atmospheric Propagation of Radiation*, Vol. 2, ERIM and SPIE Press, Bellingham, WA (1993).
- [9] Andrews, L. C., *Field Guide to Atmospheric Optics*, SPIE Press, Bellingham, WA (2004) [doi: 10.1117/3.549260].
- [10] Andrews, L. C. and Phillips, R. L., *Laser Beam Propagation through Random Media*, SPIE Press, Bellingham, WA (2005) [doi: 10.1117/3.626196].
- [11] Lukin, V. P. and Fortes, B. V., *Adaptive Beaming and Imaging in the Turbulent Atmosphere*, SPIE Press, Bellingham, WA (2002) [doi: 10.1117/3.452443].
- [12] Mayer, B., Emde, C., Buras, R., and Kylling, A., "libRadTran — library for radiative transfer," <http://www.libradtran.org>.
- [13] Nikolaeva, O. V., Bass, L. P., Germogenova, T. A., Kuznetsov, V. S., and Kokhanovsky, A. A., "Radiative transfer in horizontally and vertically inhomogeneous turbit media," *Light Scattering Reviews 2: Remote Sensing and Inverse Problems*, 295–341, Praxis Publishing, Chichester, UK (2007).

- [14] Boyd, R. W., *Radiometry and the Detection of Optical Radiation*, John Wiley & Sons, New York (1983).
- [15] Niemz, M. H., *Laser–Tissue Interactions: Fundamentals and Applications*, Springer Verlag, Berlin (2007) [doi: 10.1007/978-3-540-72192-5].
- [16] Džimbeg-Malčić, V., Barbarić-Mikočević, Ž., and Itrić, K., “Kubelka-Munk theory in describing optical properties of paper (1),” *Technical Gazette (Tehnički vjesnik)* 18(1), 117–124 (2011).
- [17] Duntley, S. Q., “The Reduction of Apparent Contrast by the Atmosphere,” *Journal of the Optical Society of America* 38(2), 179–191 (1948).
- [18] Gordon, J. I. and Duntley, S. Q., “Measuring Earth-to-Space Contrast Transmittance from Ground Stations,” *Applied Optics* 12(6), 1317–1324 (1973).
- [19] Turner, R. E., “Contrast Transmittance in Cloudy Atmospheres,” *Proc. SPIE* 305, 133–142 (1981) [doi: 10.1117/12.932706].
- [20] Justus, C. G. and Paris, M. V., “Modelling Solar Spectral Irradiance and Radiance at the Bottom and Top of a Cloudless Atmosphere,” tech. rep., School of Geophysical Sciences, Georgia Institute of Technology (1987).
- [21] Spectral Sciences Inc. and U. S. Air Force Research Laboratory, “MODTRAN,” modtran5.com.
- [22] Jursa, A. S., *Handbook of Geophysics and the Space Environment*, Vol. NTIS Document number ADA-167000, USAF Geophysics Laboratory (1985).
- [23] U.S. Environmental Protection Agency, “Characteristics of Particles,” <http://www.epa.gov/apti/bces/module3/collect/collect.htm>.
- [24] Palmer, J. M. and Grant, B. G., *The Art of Radiometry*, SPIE Press, Bellingham, WA (2009) [doi: 10.1117/3.798237].
- [25] MIL-HDBK-310, “Global Climatic Data for Developing Military Products,” Tech. Rep. MIL-HDBK-310, Department of Defense (1997).
- [26] O’Brein, S. G. and Shirkey, R. C., “Determination of Atmospheric Path Radiance: Sky-to-Ground Ratio for Wargamers,” Tech. Rep. ARL-TR-3285, Army Research Laboratory (2004).
- [27] Baker, D. J. and Pendleton Jr., W. R., “Optical Radiation from the Atmosphere,” *Proc. SPIE* 91, 50–62 (1976) [doi: 10.1117/12.955071].

- [28] Suits, G., "Radiative Transfer I, IR Technology Course," tech. rep., University of Michigan/ERIM. (unknown).
- [29] Prokes, A., "Atmospheric effects on availability of free space optics systems," *Optical Engineering* 48(6) (2009) [doi: 10.1117/1.3155431].
- [30] Kneizys, F. X., "Users Guide to LOWTRAN 7," Tech. Rep. AFGL-TR-88-0177, Air Force Systems Command, USAF (1988).
- [31] Wikipedia, "Atmospheric radiative transfer codes," http://en.wikipedia.org/wiki/Atmospheric_radiative_transfer_codes.
- [32] Ontar, "PcModWin Manual Version," tech. rep., Ontar Corporation (2001).
- [33] Kneizys, F. X., Shettle, E. P., Gallery, W. O., Chetwynd, J. H., Abreu, L. W., Selby, J. E. A., Clough, S. A., and Fenn, R. W., "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN 6," Tech. Rep. AFGL-TR-83-0187, Air Force Systems Command, USAF (1983).
- [34] Roebeling, R. A., Jolivet, D., Macke, A., Berk, L., and Feijt, A., "Inter-comparison of models for radiative transfer in clouds," *11th Conference on Atmospheric Radiation and the 11th Conference on Cloud Physics* (2002).

Problems

- 4.1 Explain what thermal radiation is, why it occurs, how it can be calculated, and how it is affected by the atmosphere. [5]
- 4.2 Describe what atmospheric aerosols are and what effect they have on optical and infrared systems operating in the spectral range 1–12 μm . Describe the effects of clean air and fog over the full extent of this spectral range. [5]
- 4.3 Provide a description of each of the following terms; explain what they are and how they work: (a) atmospheric transmittance, (b) atmospheric path radiance, (c) Rayleigh and Mie scattering, (d) discrete ordinate models, (e) molecular absorption, and (f) aerosol scattering. [12]
- 4.4 Start from first principles and derive the transmittance for a path with length R in a homogeneous medium with attenuation coefficient γ . [4]

If the spectral transmittance for a homogenous medium is known for one path length R_1 , show how the transmittance for another path length R_2 can be determined. [2]

- 4.5 Calculate and plot the blackbody exitance for temperatures of 200 K to 6000 K in the following spectral bands: (a) 0 to infinity, (b) 0.4–0.75 μm , (c) 1.5–2.5 μm , (d) 3–5 μm , (e) 3.8–4.5 μm , (f) 4.8–5.2 μm , (g) 8–12 μm , and (h) 8–14 μm . Comment on the results. [5]

Repeat the calculation but now evaluate the effect of spectral atmospheric transmittance. Use the data in the `DP03.zip` data file. [8]

- 4.6 The aerosol light scatter phase function P is the probability that light is scattered in a given direction (θ, φ) . As a probability function, $\int_{\text{sphere}} P(\theta, \varphi) d\Omega = 1$. MODTRANTM provides the user an option to use the Henyey–Greenstein scattering phase function as an approximation of Mie scattering.³³ The function is given by

$$P_{HG}(\theta) = \frac{1 - g^2}{4\pi \sqrt[3]{1 - 2g \cos \theta + g^2}}, \quad (4.37)$$

where θ is the scattering angle, and g is the asymmetry parameter; where $g = +1$ provides complete forward scattering, $g = 0$ represents isotropic scattering, and $g = -1$ provides complete back scattering. Plot this function for various values of g and compare the results with Figure 4.10. Also plot the simplified Rayleigh scattering function,³³

$$P_R(\theta) = \frac{3(1 + \cos^2 \theta)}{16\pi}. \quad (4.38)$$

Comment on your observations. [3]

The Henyey–Greenstein function is considered to be an inaccurate model of real aerosol and was replaced with a Legendre polynomial function.³⁴ Explain why. [2]

Chapter 5

Optical Detectors

The source of all light is in the eye.
Alan Wilson Watts

Cornelius J. Willers

Ricardo Augusto Tavares Santos, D.Sc.
Instituto Tecnológico de Aeronáutica, S. José dos Campos - Brazil.

Fábio Durante Pereira Alves, D.Sc.
Naval Postgraduate School, Monterey - USA.

5.1 Historical Overview

In 1800 William Herschel discovered infrared flux using a thermometer as the first infrared (IR) detector. In his experiments, a prism was used to refract sunlight. A thermometer placed just outside the red edge of the spectrum indicated a higher temperature than in the rest of the room. Early IR detectors exploited the Seebeck thermoelectric effect used in the first thermocouple devices.

The origins of modern IR detector technology can be traced to the 20th century, during World War II, when photon detectors were developed.¹ Since World War II, IR detector technology development was and continues to be primarily driven by military applications, although in the last few decades its application in civilian fields such as medicine, quality control, anti-threat systems, and industrial processes, among others, has grown substantially. This diversity of applications, as well as the advances in the semiconductor sciences and fabrication processes, leads to cost-effective devices and systems, placing IR technology in current daily life. When a new system is brought to the market today, the design specifications often consider 'dual deployment,' targeting civil and military applications.

Optical detectors are used as components in electro-optical sensor systems (see Chapter 1). The broader IR technology field concerns itself with the study of how a heated source radiates energy, how this radiation propagates through a medium, how it interacts with matter, and finally, how it is detected.

This chapter provides an introduction to IR detectors. The focus is on concepts and principles and not on specific detector materials or technologies. The classical, first-order theory presented here is suitable for basic understanding but does not cover advanced concepts or secondary effects. Starting with the physics of light absorption, the focus shifts to detector types, noise, thermal detectors, and photon detectors.

5.2 Overview of the Detection Process

The optical detection process occurs by one of two mechanisms: photon absorption and thermal energy absorption. These mechanisms lead to a number of different kinds of detectors.²⁻⁷ Table 5.1 gives an overview on IR detectors.

5.2.1 Thermal detectors

Thermal detectors respond to the heating effects of absorbed optical radiation by changing the temperature of the sensor, which causes (or induces) changes in a measurable parameter, e.g., resistance, polarization, or voltage. Thermal detectors are generally slower (the thermal processes tend to be slower) and have lower sensitivity than photon detectors. Thermal detectors' spectral response can be wide if a wideband detector coating is used. Thermal detectors were not traditionally used in high-performance electro-optical systems. Uncooled thermal detectors are becoming less expensive yet more sensitive, and find increasing use in applications previously reserved for photon detectors, such as thermal imaging cameras.

Thermal detector responsivity is defined in terms of the detector signal i_d divided by optical incident radiant flux Φ_e :

$$\mathcal{R}_{e\lambda} = \frac{i_d}{\Phi_e} = \frac{\alpha_\lambda k \Phi_e}{\Phi_e} = \alpha_\lambda k, \quad (5.1)$$

where $\mathcal{R}_{e\lambda}$ has units of [A/W] or [V/W], depending on the device, α_λ is the spectral absorption (emissivity), and k is a conversion constant. Equation (5.1) indicates that a certain amount of optical flux, expressed in [W], causes a certain output signal i_d in [A]. The device's spectral response is only a function of the spectral absorptance of the flux and not of the internal physics of the detector. Because the spectral absorptance is approx-

Table 5.1 Comparison of detector types.⁸

Detector type			Advantages	Disadvantages
Thermal	Bolometer, pyroelectric, thermopile		Light, rugged, reliable, and low cost. Room temperature operation.	Low detectivity at high frequency. Slow response (milliseconds).
Photon	Intrinsic	IV-VI PbS, PbSe, PbSnTe	Well-understood, more-stable materials. Easier to manufacture.	Very high thermal expansion coefficient. Large permittivity.
		II-VI HgCdTe	Easy bandgap tailoring. Well-developed theory and technology. Multicolor detectors.	Nonuniformity over large area. High cost in growth and processing. Surface material composition instability.
		III-V InGaAs, InAs, InSb, InAsSb	Good material & dopants. Advanced technology. Potential monolithic integration.	Heteroepitaxy with large lattice mismatch. Long wavelength cutoff limited to 7 μm (at 77 K).
	Extrinsic	Si:In, Si:Ga, Si:As, Ge:Cu, Ge:Hg	Operation at very long wavelengths. Relatively simple technology.	High thermal carrier excitation. Extremely low operating temperature.
	Free carriers	PtSi, PtSi, IrSi	Low-cost, high-yield technology. Large and close-packed arrays.	Low quantum efficiency. Low temperature operation.
	Quantum wells, Type I	GaAs/AlGaAs, InGaAs/AlGaAs	Matured material growth. Good uniformity of large area. Multicolor detectors.	High thermal carrier excitation. Complicated design and growth.
	Quantum wells, Type II	InAs/InGaSb, InAs/InAsSb	Low Auger recombination rate. Easy wavelength control.	Complicated design and growth. Sensitive to the interfaces.
	Quantum dots	InAs/GaAs, InGaAs/ InGaP, Ge/Si	Normal incidence of light. Low thermal carrier excitation.	Complicated design and growth.

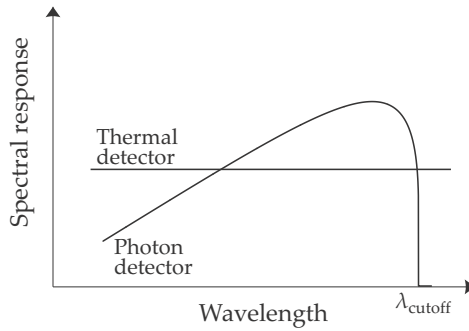


Figure 5.1 Spectral response comparison between photon and thermal detectors.

imately flat in the operational band, the responsivity curve is therefore approximately spectrally flat, as shown in Figure 5.1.

5.2.2 Photon detectors

Photon detectors rely on the absorption of a photon in a semiconductor or on an emissive surface, resulting in the release of electrons. Photon detectors have high sensitivity (good SNR) and fast response times — ideal for high-performance sensors. There is a very wide variety of photon detector technologies, the more important of which include intrinsic and extrinsic detectors, photoemissive detectors, and quantum well detectors (QWIPs).⁴ Photon detectors operating in the IR region must be cooled down to cryogenic temperatures for low-noise operation. High-performance photon detectors can be expensive.

Photon detector devices detect light by a direct interaction of the radiation with the atomic lattice of the material. This interaction produces voltage or current changes that are detected by associated circuitry or interfaces. The photon detector is a semiconductor device with a bandgap between the conduction band and the valence band. Photons with an energy exceeding the bandgap of the detector material are absorbed in the material. The energy released during absorption elevates an electron to the conduction band. Free electrons in the conduction band result in current flow, which is sensed on the detector terminals. Because photon energy increases toward shorter wavelengths, there exists a (low-energy) maximum wavelength λ_c beyond which the detector is not sensitive. The cutoff wavelength is given by

$$\lambda_c = \frac{1.24}{E_g}, \quad (5.2)$$

where λ_c is in $[\mu\text{m}]$, and the bandgap E_g is in electron-volts.

The bandgap in a semiconductor depends on complex physical parameters. One of the commonly used empirical models to calculate the semiconductor bandgap is the Varshni's approximation:

$$E_g = E_g(0) - \left(\frac{AT^2}{T+B} \right), \quad (5.3)$$

where $E_g(0)$ is the semiconductor bandgap at 0 K and A and B are Varshni material parameters.⁹⁻¹² There are also other empirical models for this purpose.¹⁰

Because the detector converts photons into electronic charges, the photon detector output signal is proportional to the number of incident photons. The responsivity can therefore be described in terms of the incoming photon flux (photons per second) as

$$\mathcal{R}_{q\lambda} = \frac{i_d}{\Phi_q} = \begin{cases} \eta q G & \text{for } \lambda \leq \lambda_c \\ 0.0 & \text{for } \lambda > \lambda_c \end{cases}, \quad (5.4)$$

where Φ_q is the incoming flux in quanta per second [q/s], i_d is the detector current [C/s], η is the spectral quantum efficiency of the detector (unitless), G is the gain of the detector material (unitless) [see Equation (5.103)], q is the electronic charge in [C], and the responsivity $\mathcal{R}_{q\lambda}$ has units of [(C/s)/(1/s)]=[C], in effect [C/quanta]. Quantum efficiency is the fraction of incident photons converted to electrons contributing to the measured signal. The gain G for photovoltaic detectors is unity because one photon creates one free electron in the junction. The gain for photoconductive detectors depends on material properties and detector design, and can be smaller or larger than unity.

From Equation (5.4) the radiant spectral responsivity of the photon detector, *defined in terms of watts*, can be found by multiplying the photon flux Φ_q with the energy per photon $Q = h\nu = hc/\lambda$, then $\Phi_e = \Phi_q hc/\lambda$ and

$$\mathcal{R}_{e\lambda} = \frac{i_d}{\Phi_e} = \frac{i_d}{\frac{hc}{\lambda}\Phi_q} = \begin{cases} \frac{\eta q \lambda G}{hc} & \text{for } \lambda \leq \lambda_c \\ 0.0 & \text{for } \lambda > \lambda_c \end{cases}, \quad (5.5)$$

where Φ_e is the incoming flux [W], i_d is the detector current [C/s] or [A], q is the electron charge [C], η is the spectral quantum efficiency of the detector (unitless), λ is wavelength in [m], G is the gain of the detector material (unitless), h is Planck's constant [J·s], c is the speed of light in a vacuum [m/s], and $\mathcal{R}_{e\lambda}$ has units of [C/J], which is the same as [A/W].

Short-wavelength photon detectors (e.g., silicon) are sensitive in the visual and NIR spectral bands. Silicon detectors are inexpensive and manufactured in large volumes for consumer electronics. Longer-wavelength

photon detectors (e.g., InSb and HgCdTe) must be cryogenically cooled to prevent thermal carrier excitation across the small semiconductor energy bandgap.

5.2.3 Normalizing responsivity

For convenience, the spectral responsivity can be written as a spectral quantity multiplied by a scaling factor $\mathcal{R} = \hat{\mathcal{R}}\tilde{\mathcal{R}}$, where $\hat{\mathcal{R}}$ is the scaling factor with units [A/W] or [V/W], and $\tilde{\mathcal{R}}$ is the spectral function with maximum value equal to unity. If a measured detector response is not available, an approximation such as Equation (D.5) can be used to calculate spectral responsivity values. The spectral response of a detector can also be calculated using the theory provided in this chapter, as shown in Figure 5.30 in Section 5.9.6.

5.2.4 Detector configurations

Detectors are found in a range of configurations, from single-element devices, to linear vectors ($N \times 1$) or ($N \times \text{few}$), to large two-dimensional arrays ($N \times M$). Detector arrays are commonly found in digital cameras, with sizes up to several tens of megapixels.

Detectors can cover a single spectral band or multiple spectral bands. Multi-color detectors can have detector elements with different colors next to each other (as in a digital camera) or behind each other (the front short-wave detector element is transparent for longer-wavelength IR radiation).

5.3 Noise

This section provides a very brief coverage of noise, as pertaining to thermal and photon detectors.^{2,7,12–16} Sensor and electronics noise can be grouped into several categories: stochastic noise processes of physical (atomic) origin, phenomenological noise ($1/f$ noise & electrical contact noise), and ‘system’ noise originating as a result of imperfections in an electronic system. There is also another noise category: originating outside the sensor but *inherent in the signal* itself.^{2,13,17–19} The different noise sources are normally uncorrelated with each other, being caused by independent processes.

Physical noise processes are inherent in any natural process or electronic component such as the detectors or transistors. Like optical aberrations and optical defects, these noise processes are mathematically and physically part of nature itself. Similarly, some noise processes can be derived

and expressed in rigorous mathematical terminology, whereas other processes are not well understood. Physical noise sources include Johnson (thermal or Nyquist) noise, shot noise, temperature-fluctuation noise, and generation–recombination (g-r) noise. Johnson noise results from random charge movement due to thermal agitation in a conductor. Shot noise arises from the statistical occurrence of discrete events, such as when charges cross a bandgap potential barrier in a semiconductor. Generation–recombination noise occurs when electron-hole pairs (with finite carrier lifetime), are generated or recombined in a semiconductor.

Defying physical explanation, *phenomenological* noise with $1/f$ fractal power spectrum is ever present in observations of physical and natural events. The noise spectrum has the form a/f^β , where a is a constant, f is electrical frequency, and β is a constant usually equal to one or two. $1/f$ noise is present in detectors, electronic circuits, flow processes such as natural rivers and traffic, biological processes, and music.²⁰

System noise sources include interference noise, fixed pattern noise in detector arrays, and microphonic noise. Interference is caused by external events injecting spurious electrical signals by capacitive, inductive, or earth loop coupling into an electronics circuit. Interference noise cannot readily be calculated from first principles, but it can be modeled as an additional noise source with magnitude $k_e NEP$, with k_e based on measurement or estimates, and NEP is defined in Section 5.3.11. Fixed pattern noise occurs in multiple element detector arrays, emanating from statistical variations between individual detector elements. Assuming a uniform illumination on all pixels in the focal plane, some pixels will provide a stronger signal than others. This difference is attributable to the nonuniformity in the detector responsivity amongst the various detector elements. Individual detector elements can have different absolute responsivity values, different spectral responsivity values, and/or nonlinear responses. Fixed pattern noise can be measured and modeled using the principles of three-dimensional noise analysis.²¹ Microphonic and triboelectric noise results from minute mechanical deformation of an electronic device or conductors, causing signal generation in piezoelectric or triboelectric materials, or variation in device capacitance.

5.3.1 Noise power spectral density

Noise in a detector or electronic circuit can be expressed as a voltage v or a current i , the choice of which depends on the impedance levels in the design or the nature of the signal. Both forms are used in this book. In its more-fundamental form, noise is expressed as a power spectral density

(PSD), with units of $[W/Hz]$, $[A^2/Hz]$, or $[V^2/Hz]$. It quantifies the content of the noise signal over a small bandwidth (say, 1 Hz) at a particular frequency. In this case the spectral domain is not optical wavelength but the temporal variation of a signal expressed in units of $[Hz]$ (cycles per second). The noise density can also be expressed in volts or current with units of $[A/\sqrt{Hz}]$ or $[V/\sqrt{Hz}]$.

The PSD of a time signal $f(t)$ is given by

$$S_f(\omega) = \left\langle \lim_{T \rightarrow \infty} \frac{|F_T(\omega; \beta)|^2}{2T} \right\rangle, \quad (5.6)$$

where the expected value $\langle \rangle$ is taken over the ensemble of signals β (not time t), and $F_T(\omega; \beta)$ is the Fourier transform of the time signal $f(t)$:

$$F_T(\omega; \beta) = \int_{-T}^T f(t; \beta) e^{-i\omega t} dt. \quad (5.7)$$

In effect this means that several (as in a very large number) infinitely long sequences of the time signal $f(t)$ are transformed by the Fourier transformation and multiplied with their complex conjugates. The ensemble PSD is then obtained by averaging the individual spectra.

White noise has a constant PSD at all frequencies (spectrally flat). In contrast, band-limited noise may have an arbitrary PSD, depending primarily on the frequency response of the electronic filter used to define the bandwidth. $1/f$ noise has constant noise power per frequency octave (frequency ratio of 2).

The integral of the PSD over all temporal frequencies yields the total power in the signal:

$$P = \int_{-\infty}^{\infty} S_f(\omega) d\omega. \quad (5.8)$$

Noise can therefore also be expressed in an integrated form with units of $[W]$, $[A^2]$, or $[V^2]$. Depending on the context and the units used, it should be clear whether the spectral or integral values are used.

Note that the power discussed here is electrical power in an electronic circuit, which is not the same as optical flux power. They share the same unit and fundamental concept of power but have different contextual meanings.

5.3.2 Johnson noise

Johnson noise is also known as Nyquist or thermal noise.^{2,22} It is caused by the random Brownian motion of carriers in a conductor with nonzero resis-

tance at nonzero temperatures. Johnson noise is only generated in the dissipative real component of complex impedances, present in almost all electronic components. Johnson noise can be considered as a one-dimensional form of blackbody radiation because its derivation is based on the density of states (see also Sections 3.1 and 5.5.3).²³ Johnson noise power spectral density (for positive frequencies) is given by

$$v_n^2 = 4kTR \left(\frac{x}{e^x - 1} \right) \quad \text{and} \quad (5.9)$$

$$i_n^2 = \frac{4kT}{R} \left(\frac{x}{e^x - 1} \right), \quad (5.10)$$

where v_n and i_n are the rms noise voltage and current power spectral densities in $[V^2/\text{Hz}]$ and $[A^2/\text{Hz}]$, respectively. k is the Boltzmann constant, T is the temperature in $[K]$ of the resistive element with value R in $[\Omega]$, and $x = hf/(kT)$, where f is the electrical frequency. The term

$$\frac{x}{e^x - 1} \quad (5.11)$$

describes the frequency spectrum of the noise — note its similarity with the Planck-law formulation. hf is the energy of a particle at frequency f , whereas kT is the kinetic energy of a particle at temperature T . Because $e^x = 1 + x + x^2/2! + x^3/3! + \dots$, it follows that for small x the term has unity value. Investigation shows that the Johnson noise spectrum is flat up to frequencies of the order of 10^{12} Hz.

The probability density function of unfiltered Johnson noise is Gaussian with variance $i_n^2 \Delta f$ or $v_n^2 \Delta f$, where Δf is the electronic noise bandwidth.

5.3.3 Shot noise

Shot noise is the disturbance caused by a constant presence of random discrete events. Examples of such events are the flow of carriers across a potential barrier, the detection of single photons in very low light level observations, or hail stones on a sheet metal roof. Shot noise is generated in all photovoltaic and photoemissive detectors. It is not found in photoconductive detectors. Being the accumulation of discrete events, the PDF of shot noise is Poisson. The shot noise equation is equally valid for photons, electrons, or holes. Note that shot noise is independent of temperature.

At low frequencies (and for all practical purposes) shot noise PSD is given by

$$i_n^2 = 2qI, \quad (5.12)$$

where i^2 has units of $[A^2/Hz]$. I is the average current in $[A]$ or $[q/s]$, and q is the charge on an electron $[C]$. This equation is easily derived from the Poisson statistics, which states that the variance in events is equal to the mean of the number of events $S_q(0) = \sigma_q^2 = a_q$, where the subscript q denotes quanta (electrons or photons). If q is the charge of one electron, it follows that

$$S_i(0) = q^2 S_q(0) = 2q^2 \sigma_q^2 = 2q^2 a = 2qI, \quad (5.13)$$

where $S_i(0)$ and $S_q(0)$ are the low-frequency PSDs of the current and quantum rates, respectively. The factor 2 is introduced to allow for positive bandwidths only, i.e., integrating S_i as $\int_0^{\omega_b} S_i(\omega) d\omega$ instead of $\int_{-\omega_b}^{\omega_b} S_i(\omega) d\omega$.

For large numbers of events, the Poisson distribution approximates a Gaussian distribution. This means that shot noise has a Gaussian distribution for large photon-flux rates. The distribution is not Gaussian when small event numbers are considered, such as low light level at visual or ultraviolet wavelengths.

The autocorrelation between consecutive events is the Dirac delta function (i.e., no correlation), hence the PSD of shot noise is spectrally flat.

5.3.4 Generation–recombination noise

Generation–recombination noise is caused by the creation and recombination of electron-hole pairs in a photoconductive detector. The incoming photon flux creates free carriers that increase the conductance of the detector. However, the finite lifetime of these carriers means that they recombine, resulting in a decrease of the conductance. The net effect is that the continuous generation and recombination of free carriers result in noise in the signal. In photoconductive detectors, the g-r noise is the primary noise contributor at medium-to-high frequencies. Generation–recombination noise is not found in most photovoltaic detectors because the carriers are quickly swept out of the junction before recombination. Some materials do have g-r noise in photovoltaic devices.

The generation of free carriers can be caused by two processes: thermally excited carriers or optically excited carriers. Thermally excited carriers are created by virtue of the energy levels in the material due to its operating temperature. At higher temperatures, more thermally excited carriers are generated (kT). The rms g-r noise current in a photon detector is given by²

$$i_{gr} = 2qG \sqrt{\eta E_q A \Delta f + g_{th} A_d \Delta f l_x}, \quad (5.14)$$

where q is the charge on an electron in $[C]$, G is the photoconductive gain

(unitless), η is the detector quantum efficiency (unitless), E_q is the photon flux on the detector in $[q/(s \cdot m^2)]$, A_d is the area of the detector in $[m^2]$, Δf is the noise equivalent bandwidth in $[Hz]$, g_{th} is the rate of thermal carrier generation, and l_x is the thickness in the optical propagation direction in $[m]$.

The g-r noise power spectral density has the general form

$$S \propto \frac{t^2}{1 + f^2 t^2} \quad (5.15)$$

where t is the carrier lifetime in $[s]$, and f is the electrical frequency. Note that the carrier lifetime introduces a pole in the transfer function, resulting in a roll-off at higher frequencies. The high-frequency roll-off is not as high as for shot or thermal noise because the carrier lifetimes are much longer than the brief events associated with shot noise and thermal noise.

5.3.5 $1/f$ noise

$1/f$ noise is a term loosely applied to a broad range of noise sources, most of which are not well understood. $1/f$ noise is found in virtually all detectors, music, stock prices, rainfall records, literature, and other fractal and natural signals. The very deep origins of $1/f$ noises are not well understood in most cases. Electronic manifestations of $1/f$ noise include contact noise, flicker noise, and excess noise. $1/f$ noise causes detector noise at frequencies up to several kHz. The noise PSD is given by

$$i_n^2 = \frac{k_1 I^\alpha}{f^\beta}, \quad (5.16)$$

where i_n^2 has units $[A^2/Hz]$, k_1 is a constant, I is the (mostly DC) current through the device, $1.25 < \alpha < 4$ (usually 2), and $0.8 < \beta < 3$ (usually 2).

5.3.6 Temperature-fluctuation noise

Even in thermal equilibrium, the particle movement in an object causes temperature fluctuations around the mean temperature. The temperature distribution is white noise Gaussian with a mean value of T and variance²⁴

$$\overline{(\Delta T)^2} = \frac{kT^2}{C}, \quad (5.17)$$

where k is Boltzmann's constant, T is the mean temperature in $[K]$, and C is the heat capacity of the object in $[J/K]$. It can be shown that the temperature-fluctuation noise PSD is^{12,24–26}

$$(\Delta T)_n^2 = \frac{4kT^2}{\sqrt{G^2 + (\omega C)^2}} \quad (5.18)$$

in $[\text{K}^2/\text{Hz}]$, from which the corresponding noise flux Φ can be determined as^{24,25}

$$\Phi = 4kGT^2 \quad (5.19)$$

in $[\text{K}^2/\text{Hz}]$ for low frequencies, where k is the Boltzmann constant, T is the object's temperature, and Δf is the noise equivalent bandwidth. The meaning of G and C is discussed in Section 5.4.2.

5.3.7 Interface electronics noise

The detector electronics interface also contributes noise.^{14,15,27} This noise is inherent in the electronic components (shot noise or Johnson noise) or can be a result of processing (amplitude digitization/sampling noise).

5.3.8 Noise considerations in imaging systems

A subtle cause of noise in an imaging system results from the spatial sampling (aliasing) of small objects in an image — this effect is not a noise source, but it manifests itself in the form of semi-random variations in signal strength. Section B.4.1 describes this effect in a case-study context. The performance of imaging systems, including noise effects, are well documented.^{28–30} Noise in imaging systems is best analyzed in the context of the three-dimensional noise model.^{21,28,31}

5.3.9 Signal flux fluctuation noise

Optical flux generation is a Poisson process, which carries with the average value an inherent noise variance equal to the mean flux level. The shot noise fluctuation is smaller at IR wavelengths but becomes significant at very low flux levels for visual or UV light. Fewer photons are required in the visual or UV spectral bands than in the IR spectral bands because the photons have higher energy *per photon* at shorter wavelengths. The noise inherent in the signal sets a minimum detectable signal level. The following derivation determines this minimum signal level under the assumption that there are no other signal or noise sources.

An optical flux of Φ_p [q/s] results in a current of $I = \eta q \Phi_p$ in a detector. This current causes shot noise (see Section 5.3.3) with a magnitude

$$i_n^2 = 2qI\Delta f = 2q^2\eta\Phi_p\Delta f. \quad (5.20)$$

$\Phi_e = hc\Phi_p/\lambda$, hence the SNR becomes

$$\text{SNR} = \sqrt{\frac{\eta\lambda\Phi_e}{2hc\Delta f}}. \quad (5.21)$$

The noise equivalent signal power (signal where $\text{SNR} = 1$) then follows as

$$NEP = \frac{2hc\Delta f(1)^2}{\eta\lambda}. \quad (5.22)$$

Equation (5.22) states that for a given wavelength λ and noise bandwidth Δf , the noise equivalent power *due to the signal fluctuation* is determined by the detector quantum efficiency η .

It can likewise be shown that, for a wideband photon detector, the noise equivalent power due to signal fluctuation is

$$\Phi = \frac{2\Delta f \int_{\nu_0}^{\infty} M_{ev}(T_s)d\nu}{\eta \int_{\nu_0}^{\infty} \frac{M_{ev}(T_s)d\nu}{h\nu}}, \quad (5.23)$$

where Φ has units of [W], Δf is the noise bandwidth, η is the detector quantum efficiency (assumed spectrally constant), ν is frequency in [Hz], M_e is the thermal exitance in [W/m^2] from a blackbody at a temperature of T_s , k is Boltzmann's constant, h is Planck's constant, and c is the speed of light. The integration starts at $\nu_0 = c/\lambda_0$, which is the lowest frequency that the sensor can detect.

5.3.10 Background flux fluctuation noise

The background flux also causes noise in the detector signal. If the sensor is limited by the noise caused by the background, the sensor is said to be operating at 'BLIP' (background-limited performance). The principle used to determine the BLIP limit is the same as was used for signal-limited performance, except that in this case the limit is set by the noise caused by the background.

For a monochromatic source at frequency ν , the minimum detectable power ($\text{SNR} = 1$) of an open detector against a thermal radiation background is¹⁹

$$NEP = \frac{h\nu}{\eta(\nu)} \sqrt{\left[2AB \int_{\nu_0}^{\infty} \frac{\eta(\nu)2\pi\nu^2 \exp(h\nu/kT_b)d\nu}{c^2[\exp(h\nu/kT_b) - 1]^2} \right]}, \quad (5.24)$$

where Δf is the bandwidth, A is detector area, $\eta(\nu)$ is the detection quantum efficiency, ν is the frequency, T_b is the background temperature, k is Boltzmann's constant, h is Planck's constant, and c is the speed of light.

5.3.11 Detector noise equivalent power and detectivity

The electronic noise in a detector, normally expressed in terms of voltage or current at the output of the detector or electronics, can be recalculated as

an equivalent optical flux power in the detector. This optical noise power does not have the same units as the electronic noise power (discussed in Section 5.3) and is called the noise equivalent power (NEP). The NEP is the optical signal, in [W] or [q/s], required to give the same electrical signal as the noise signal. From its definition, spectral NEP is given by

$$NEP_{\lambda} = \frac{i_d \sqrt{\Delta f}}{\mathcal{R}_{\lambda}}, \quad (5.25)$$

where i_d is the detector noise current density in [A/ $\sqrt{\text{Hz}}$], Δf is the noise equivalent bandwidth in [Hz], and \mathcal{R}_{λ} is the spectral detector responsivity in [A/W]. NEP is expressed in units of [W].

NEP is the noise for a particular detector device. Most noise source contributions scale with detector area, and it is convenient to derive a more-universal measure of detector performance specific detectivity (D^*), which is normalized with respect to area and frequency. D^* is normally given in units of [cm $\cdot\sqrt{\text{Hz}}$ /W], not [m $\cdot\sqrt{\text{Hz}}$ /W]. The detector's D^* and NEP are related by

$$D_{\lambda}^* = \frac{\sqrt{\Delta f A_d}}{NEP_{\lambda}} = \frac{\mathcal{R}_{\lambda} \sqrt{\Delta f A_d}}{i_d \sqrt{\Delta f}}, \quad (5.26)$$

where A_d is the detector area in [m²]. The D^* is a more-fundamental property of the noise processes in detector materials and is independent of the detector geometry. IR detectors are normally specified in terms of D^* at some background and source temperature. The noise of visual and NIR detectors (e.g., silicon) are generally specified in terms of NEP.

NEP and detectivity can be expressed in spectral terms, NEP_{λ} , or in wideband terms, averaged over a spectral range. Similar to the spectral values described above, the wideband values can be defined. Wideband NEP is defined as

$$NEP_{\text{eff}} = \frac{i_d \sqrt{\Delta f}}{\mathcal{R}_{\text{eff}}}, \quad (5.27)$$

where i_d is the detector noise current density in [A/ $\sqrt{\text{Hz}}$], Δf is the noise equivalent bandwidth of the system, and \mathcal{R}_{eff} is the effective detector responsivity in [A/W], defined by (see Section 7.2.2)

$$\mathcal{R}_{\text{eff}} = \frac{\int_0^{\infty} \mathcal{R}_{\lambda} \tau_a M_{\lambda} d\lambda}{\int_0^{\infty} \tau_a M_{\lambda} d\lambda}, \quad (5.28)$$

where \mathcal{R}_{λ} is the detector spectral responsivity, τ_a is the spectral transmittance of the atmosphere or filters, and M_{λ} is the reference or calibration source spectral exitance.

Wideband D^* and NEP are related by

$$D_{\text{eff}}^* = \frac{\sqrt{\Delta f A_d}}{\text{NEP}_{\text{eff}}} = \frac{\mathcal{R}_{\text{eff}} \sqrt{\Delta f A_d}}{i_d \sqrt{\Delta f}}. \quad (5.29)$$

5.3.12 Combining power spectral densities

One or more different noise sources can be present in a sensor system. The different noise sources are mostly uncorrelated, originating from different components each with individual, statistically independent processes. It can be shown that *for uncorrelated sources*, noise power adds linearly. The total noise is then given by

$$i_{\text{eff}} = \sqrt{\sum_0^N i_n^2}, \quad (5.30)$$

where there are N noise sources i_n . The individual noise sources in this equation can be either spectral noise [$\text{A}\sqrt{\text{Hz}}$] or integrated wideband noise [A].

A little care must be taken with noise expressed in terms of NEP. In this case optical noise (NEP) corresponds to the signal itself, not the noise it ‘mimics.’ Hence, when NEP from different noises are combined, the NEP components must add in the square of NEP:

$$\text{NEP}_{\text{eff}} = \sqrt{\sum_0^N \text{NEP}_n^2}. \quad (5.31)$$

D^* likewise also adds in quadrature,

$$\frac{1}{D_{\text{eff}}^*} = \sqrt{\sum_0^N \frac{1}{(D_n^*)^2}}. \quad (5.32)$$

5.3.13 Noise equivalent bandwidth

Given a noise PSD $S(f)$ as input to a filter with voltage gain A_v , the noise equivalent bandwidth of the filter is defined as

$$\Delta f = \frac{1}{\max(A_v^2 S)} \int_0^\infty A_v^2(f) S(f) df, \quad (5.33)$$

where f is electrical frequency [Hz], $A_v(f)$ is the voltage frequency response of the filter or electronics circuit (unitless), $S(f)$ is the noise PSD in $[V^2/\text{Hz}]$, and $\max(A_v^2 S)$ is the maximum value of $A_v^2(f)S(f)$. It is convenient to define a constant k_n as the ratio of noise equivalent bandwidth to -3 dB bandwidth of the filter.

In the general case where the noise PSD is not flat, it follows that the noise equivalent bandwidth of a filter is a function of not only the filter but also the noise spectrum. If the noise PSD has a significant $1/f$ noise component, it means that the filter must suppress at least the lower frequencies otherwise the maximum is not defined. See also Section 7.2.3.

5.3.14 Time-bandwidth product

The time-bandwidth product relates the temporal width of a pulse to the required electrical bandwidth to achieve a specific design objective. Typical objectives include (1) the need to maintain the signal shape or (2) to achieve the best SNR, irrespective of the resulting pulse shape. The numerical value for the time-bandwidth product depends on the nature of the signal, the nature of the noise, and the required objective. The value is best determined uniquely for every different application. There are several different definitions, often leading to confusion. The exact definition of the pulse width and electronic bandwidth must therefore accompany the numerical value. A commonly used definition for temporal pulse width is the width of the pulse at half of its peak value. Filter bandwidths are likewise often defined in terms of the full-width-half-maximum (FWHM) bandwidth (also known as the -3 dB frequency bandwidth). There is no general symbol defined for the time-bandwidth product; in this book the symbol k_f is used for this purpose.

One design approach matches the filter to the shape of the signal, i.e., the filter frequency response is the Fourier transform of the signal temporal shape. King³² showed that for matched filters typical values for the time-bandwidth product constant are approximately 1 for a $\sin(x)/x$ pulse, 1.21 for a rectangular pulse, 1.64 for a half-sine pulse, and 1.77 for a trapezium-shaped pulse. The time-bandwidth factor is an approximation, depending on a number of assumptions and conventions. Its use here is to obtain order-of-magnitude values. Typical values range from 0.5 to 2. It may be convenient to consider the 'generic' value of this product to be $\pi/2$ because it sometimes simplifies calculations.

5.4 Thermal Detectors

5.4.1 Principle of operation

The scope of thermal detector technology includes a large number of different physical mechanisms, resulting in a wide variety of measurable characteristics. Common to all these mechanisms is the underlying principle that the absorbed heat increases the temperature of the device (hence the name thermal), which is observed in the change of some observable property of the device. The two major groupings include thermoelectric transducer effects (Seebeck effect and pyroelectric effect) and parametric transducers where the device temperature modulates an electric signal (resistive bolometers, Golay cell, and p-n diodes).^{25,33}

The Peltier–Seebeck effect (discovered independently by Peltier, Seebeck and Thomson) is the bidirectional conversion between temperature and voltage. This effect is exploited in thermocouple devices and Peltier coolers used to cool detectors and mini-fridges. In a pyroelectric device, temperature variations result in dielectric polarization changes in the material. Pyroelectric detectors are commonly used in IR movement detectors in security applications. In bolometer detectors the temperature change results in a change in the device's resistance. Nanotechnology bolometers are used in low-cost thermal imaging applications.

In addition to the types mentioned earlier, there are several other effects also exploited in thermal detectors.²⁵ None of these effects require the use of small-bandgap semiconductor materials, thereby not requiring material cool-down for long-wavelength operation. Some of the thermal detectors remain sensitive to temperature effects, such as pyroelectric detectors that lose polarization above the Curie temperature.

A common requirement for all thermal detectors is that the sensing element must be thermally isolated from ambient temperature structures in order to allow minute temperature changes in the sensing element. A conceptual model of a thermal detector is shown in Figure 5.2. The detector-element thermal balance is affected by three heat-flow paths: (a) the incident flux from the object (target), (b) thermally radiated flux from the detector element, and (c) heat conducted from the detector element to the device's substrate. Thermal detector performance optimization entails the careful optimization of the heat balance equation.

Modern thermal detectors employ elements with very small thermal mass (heat capacity) compared to the surface area of the device. For a given amount of absorbed energy, the temperature rise is maximized. Likewise,

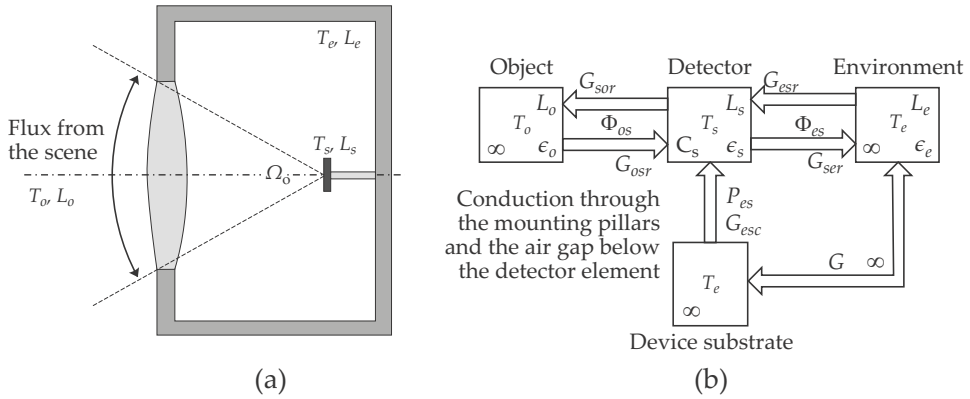


Figure 5.2 Conceptual model for thermal detector: (a) physical layout and (b) flux flow model.

because the radiating area is large compared to the thermal mass of the detecting element, the detector quickly cools down once the incident energy source is removed.

5.4.2 Thermal detector responsivity

In Figure 5.2 the detector element with heat capacity C_s in [J/K] is at a temperature T_s and radiates with radiance L_s into a full spherical environment. The environment is at a temperature T_e , radiates with radiance L_e , and has an infinite heat capacity (it can source or sink an infinite amount of energy without changing temperature). The target object is at a temperature T_o , radiates with radiance L_o , and has an infinite heat capacity. The detector element is fixed by mounting posts to the environment (the readout electronics interface chip). The mounting posts conduct heat P_{es} with conductance G_{esc} in [W/K] from the detector element to the environment. The radiative flux exchange between the object and the detector element is indicated by Φ_{os} . The radiative flux exchange between the detector element and the environment is indicated by Φ_{es} . The mounting posts between the detector element and the environment have a collective heat conductance G_{esc} . In this analysis the detector element is considered a thin disk with area A_d . Under thermal equilibrium the net inflow of power on the detector element is zero, $\Phi_{os} + \Phi_{es} + P_{es} = 0$, where inflowing power is positive:

$$0 = \int_0^\infty A_d \Omega_o (\alpha_{s\lambda} L_{o\lambda} - \alpha_{o\lambda} L_{s\lambda}) d\lambda + \int_0^\infty A_d (2\pi - \Omega_o) (\alpha_{s\lambda} L_{e\lambda} - \alpha_{e\lambda} L_{s\lambda}) d\lambda + G_{esc} (T_e - T_s), \quad (5.34)$$

where Ω_o is the optics FOV, α_e is the environment absorptance (emissivity), α_s is the detector-element absorptance (emissivity), and α_o is the object's absorptance (emissivity). Equation (5.34) applies to the total flux over all wavelengths. Each of the three radiative sources has its own spectral emissivity, requiring a detailed spectral radiometry analysis to find the solution. The following analysis assumes constant spectral absorption and emissivity by applying Kirchhoff's law and setting all values equal to a scalar value $\alpha_o = \alpha_e = \alpha_s = \epsilon = \epsilon_o = \epsilon_e = \epsilon_s = \epsilon$. Next, perform the spectral integrals, resulting in:

$$0 = A_d \epsilon [\Omega_o (L_o - L_s) + (2\pi - \Omega_o) (L_e - L_s)] + G_{esc} (T_e - T_s). \quad (5.35)$$

The two detector heat-loss mechanisms are radiation heat loss and heat conduction to the environment. Consider these two cases separately. Case 1: $L_e = L_s$, no heat loss via radiation to the environment. Then $A_d \epsilon \Omega_o L_o = G_{esc} (T_s - T_e)$, where the object radiance L_o causes a detector-element temperature T_s . Case 2: $G_{esc} = 0$, isolated detector element, with no physical contact with the environment. Then $\Omega_o L_o = 2\pi L_s - (2\pi - \Omega_o) L_e$, where object radiance L_o raises the detector-element temperature such that it radiates at L_s . In general, both radiation and conduction to the environment take place, the relative ratio of which depends on the heat capacity and spectral emissivity values of the three components in this system.

By the Stefan–Boltzmann law, Equation (3.19), the flux radiated (or lost) by a Lambertian object over all wavelengths is

$$\Phi_{SB} = \frac{A \sigma_e T^4}{\pi}, \quad (5.36)$$

where A is the radiating surface area, ϵ in this case is the effective hemispherical emissivity, and T is the temperature. The temperature derivative of the wideband flux is

$$\frac{d\Phi_{SB}}{dT} = \frac{4A\sigma_e T^3}{\pi} = G_r(T) \quad (5.37)$$

with units [W/K], which, by definition, is thermal conductance. $G_r(T)$ can be interpreted that a thermal radiator loses flux by a 'conductance' given by $4A\sigma_e T^3/\pi$ — this is not a physical conductance, but it has the equivalent effect. This 'conductance' varies with temperature.

The derivative of Equation (5.35) with respect to temperature is (keeping in mind that $dL_e/dT = 0$ because T_e is constant)

$$\begin{aligned} A_d \epsilon \Omega_o \frac{dL_o}{dT} &= A_d \epsilon 2\pi \frac{dL_s}{dT} + G_{esc}, \text{ hence} \\ \epsilon \frac{d\Phi_o}{dT} &= \epsilon \frac{d\Phi_s}{dT} + G_{esc} = \epsilon G_r(T_s) + G_{esc}, \text{ and finally} \\ \epsilon \frac{\Delta\Phi}{\Delta T} &= G = \epsilon G_r(T_s) + G_{esc}, \end{aligned} \quad (5.38)$$

which defines a detector thermal conductance for small changes in Φ_o incident on the detector.

In the closed system in Figure 5.2(a), the absorbed flux has two effects: a change in the detector-element temperature $\Phi dt = d(\Delta T) C$ as well as a heat loss through conduction to the substrate $P = G\Delta T$. Thus, the temperature of the detector element is given by the solution of the differential equation²⁵

$$C \frac{d(\Delta T)}{dt} + G\Delta T = \epsilon(\Delta\Phi)(t), \quad (5.39)$$

where G is given by Equation (5.38). Assuming a sinusoidal input signal $\Delta\Phi(t) = \Delta\Phi e^{i\omega t}$, the responsivity of the detector can be derived as being of the general form¹²

$$\Delta T = \Delta T_0 \exp(t/\tau_\theta) + \frac{\epsilon(\Delta\Phi)e^{i\omega t}}{G + i\omega C}, \quad (5.40)$$

where $\tau_\theta = C/G$ is the thermal time constant of the detector, and T_0 is the initial state of the detector. The transient exponential term becomes zero for large t . The magnitude of ΔT then becomes

$$\begin{aligned} \Delta T &= \frac{\epsilon(\Delta\Phi)}{\sqrt{G^2 + (\omega C)^2}} \\ &= \frac{\epsilon(\Delta\Phi)}{G\sqrt{1 + (\omega\tau_\theta)^2}}. \end{aligned} \quad (5.41)$$

The responsivity is then

$$\mathcal{R} = \left(\frac{\Delta T}{\Delta\Phi} \right) \left(\frac{i_d}{\Delta T} \right) = \frac{g}{G\sqrt{1 + (\omega\tau_\theta)^2}}, \quad (5.42)$$

where \mathcal{R} is the responsivity in [A/W], and g depends on the conversion mechanism for the type of thermal detector. Responsivity can be similarly defined in terms of voltage output.

One technique to improve the frequency response of the detector is to increase the detector conductance G . This would, however, result in a reduced responsivity, as shown in Equation (5.42). The better way to improve the frequency response is to reduce the detector-element heat capacity C . The heat capacity is $C = c\rho V$, where $c = dC/dm$ is the detector material specific heat in [J/(g·K)], ρ is the material density in [g/m³], and V is the detector-element volume in [m³]. The material properties c and ρ are fixed; the only design freedom is the volume. The detector area must be maximized; therefore the detector-element thickness must be minimized to reduce the element's heat capacity — this can be done with no other detrimental effect on performance. The only requirement on the detector-element thickness is to achieve mechanical stability and rigidity.

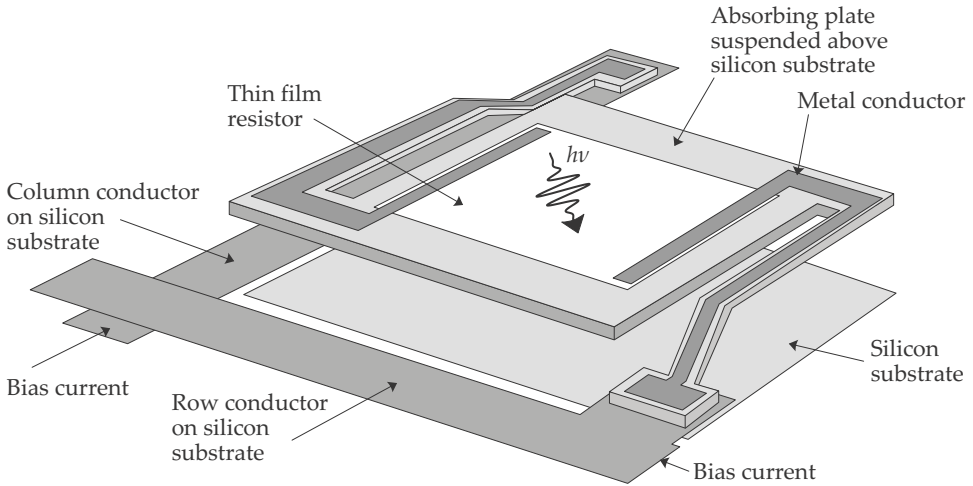


Figure 5.3 Resistive bolometer construction (adapted³⁴).

5.4.3 Resistive bolometer

The resistive bolometer senses a change in electrical resistance of the device when the device temperature changes as a result of changing absorbed radiant energy. The bolometer consists of a thin, absorbent, metallic or semiconductor layer on a structure that is thermally isolated from the substrate material (low conductance). Different materials, ranging from metals to semiconductors, can be used for the resistive element in these detectors. The structure is designed to maximize absorption and the associated temperature increase but minimize heat loss to the substrate. Microbolometer detector elements are constructed using nanotechnology processes, achieving elements with very good thermal performance.^{25,33,35–37} Figure 5.3 shows the construction of one type of microbolometer detector, used in modern staring array detectors. The readout electronics is located underneath each detector element.

The metallic element bolometer has a positive temperature coefficient of resistance (resistance increases when temperature increases). The temperature coefficient of resistance α_B in $[\text{K}^{-1}]$ is defined as

$$\alpha_B = \frac{1}{R_B} \frac{dR_B}{dT_s}, \quad (5.43)$$

where R_B is the bolometer resistance. The detector-element resistance is given by the parametric equation

$$R_B(T) = R_{B0} [1 + \alpha_B(T - T_0)], \quad (5.44)$$

where R_{B0} is the resistance at temperature T_0 .

If the thermal energy kT of an electron exceeds the bandgap in a semiconductor material it could be excited into the conduction band (see Section 5.5). The resulting holes in the valence band and electrons in the conduction band then contribute to electrical conduction. Semiconductor materials have negative temperature coefficients of resistance:

$$R_B(T) = R_0 \sqrt[n]{T} \exp(b/T), \quad (5.45)$$

where R_0 is a reference resistance, T is the temperature, and n and b are constants. Rogalski¹² states that $n = -3/2$, and b is a material constant. Budzier²⁴ and others state that $n = 0$ and $b = E_g/(2k)$, where E_g is the material bandgap and k is Boltzmann's constant (see Section 5.5). Note that E_g itself is also a function of temperature. The temperature coefficient of resistance is then

$$\alpha_B = -\frac{b}{T^2}. \quad (5.46)$$

The output from the bolometer is proportional to the input flux, and the device is therefore able to respond to static radiant flux input conditions, i.e., when viewing a scene with no moving objects. This means that thermal cameras with bolometer devices do not need optical choppers to operate. Nonuniformity correction may require that the detector observes one or more uniform temperature reference sources during calibration.

The responsivity of the bolometer detector element is given by^{12,24,25}

$$\mathcal{R} = \frac{\alpha_B \epsilon I_b R_b}{G \sqrt{1 + (\omega \tau_\theta)^2}}, \quad (5.47)$$

where α_B is the temperature coefficient of resistance, ϵ is the effective wideband surface absorption (emissivity), I_b is the constant bias current through the device, R_b is the device electrical resistance, and G is the heat conduction coefficient, given by Equation (5.38).

Modern microbolometer structures are based on the 'monolithic' design concept by Honeywell, making use of silicon micro-machining.^{25,35} The sensing element is a 0.5–3- μm thin membrane of silicon nitride (Si_3N_4), supported by two pillar structures or posts, covered with a thin film of the resistive material, usually a material such as vanadium oxide (VO_x). The posts perform a structural function to support the membrane in free space but also act as the substrate for a conducting metal layer that connects the resistive element on the membrane to the underlying read-out electronics. The silicon-nitride membrane has gold on the back (bottom) side to reflect energy from the substrate back into the device.

The whole structure is formed by first depositing a layer of silicon dioxide on the silicon chip. The membrane and posts are then grown on

top of the SiO_2 , after which the SiO_2 is etched away, leaving the membrane only, supported by the posts. The vanadium oxide film is formed by sputtering a layer of mixed oxides on the membrane. The mixed-oxide layer provides a more-stable detector than would pure oxides. The mixed-oxide temperature coefficient of resistance α is of the order of -2 to $-4\ \%/^{\circ}\text{C}$. Other materials can also be used, and in one case, a forward-biased p-n diode is used as the temperature sensor.

Microbolometers are available in two-dimensional arrays. The individual element sizes can be as small as $17\ \mu\text{m}$, but common sizes range from 25 – $50\ \mu\text{m}$. Some devices employ a two-story construction to increase the fill factor. The array structure is built on top of a silicon read-out electronics interface chip (ROIC).

Microbolometers have achieved relatively high performance with thermal imaging cameras demonstrating noise equivalent temperature difference of $0.1\ \text{K}$ using $f/1$ optics.³⁵ Best performance is achieved if the detector is encapsulated in a vacuum (pressure less than $1\ \text{mbar}$) to reduce thermal conductance through the air, in the small gap between the bottom of the detector element and the substrate underneath.

The bolometer device has the following noise sources:³⁵ (1) Johnson noise (see Section 5.3.2), (2) $1/f$ noise (see Section 5.3.5), (3) temperature-fluctuation noise (see Section 5.3.6), and (4) read-out electronics interface circuit noise (see Section 5.3.7). These noise sources are uncorrelated and add in quadrature (see Section 5.3.12).

5.4.4 Pyroelectric detector

The pyroelectric effect^{25,33,38} is found in many different materials, but ferroelectric materials are more commonly used. The ferroelectric detector senses the changes in the electrical polarization of the material resulting from temperature changes. The output from the ferroelectric detector is proportional to the rate of change in input flux — if the scene flux does not change, the signal disappears. Sensors with pyroelectric detectors rely on movement in the scene or require a device to ‘chop’ the incident signal, alternating the scene flux with a reference flux.

The ferroelectric effect is found in materials such as barium strontium titanate, strontium barium niobate, lithium tantalate, and lead titanite. These materials have spontaneous internal polarization, measured as a voltage on electrodes placed on opposite sides of the bulk of the material; this forms a capacitor with the sensor material as the dielectric. At a constant temperature, the polarization is equalized by mobile charges on the

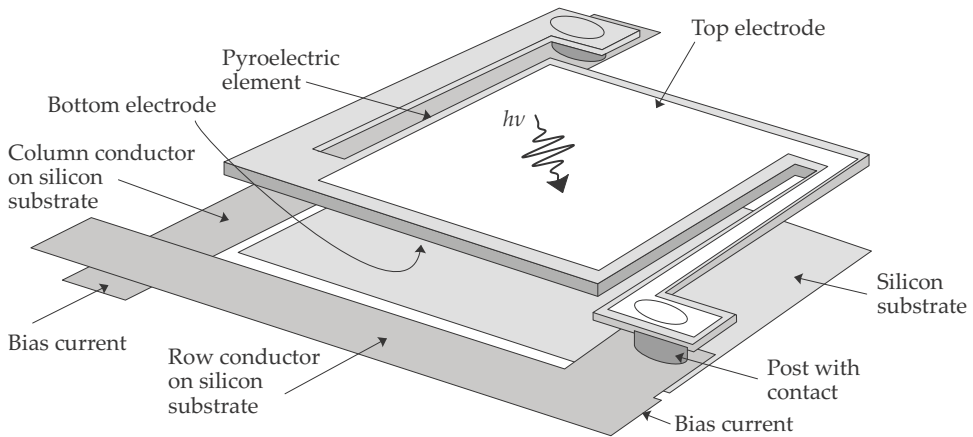


Figure 5.4 Pyroelectric detector structure.²⁵

surface of the material. As the bulk temperature changes, the polarization of the material changes, resulting in a temporary change in charges at the surface and leads to a minute current flowing to restore the charges. The current flow taking place across the detector's capacitive structure results in a voltage change. The voltage changes are only observed when the temperature changes, hence the element temperature must be time-modulated to produce an output from a static scene. This can be done by chopping the signal mechanically or by moving the image across the detector.

The ferroelectric effect is maximum just below the Curie temperature, and the detector is often operated at this point. Care must be taken not to exceed the Curie temperature, as the device will lose its polarization above this temperature. Losing polarization is a temporary effect, but the device must be re-poled (application of a high field strength at high temperature).

One approach to pyroelectric detector layout is shown in Figure 5.4. The structure employs micro-machined silicon technology to construct an isolated element, supported by two posts. The dielectric material spans the full detector area, with one electrode on the top side, and the other electrode on the bottom side. This forms the device's capacitor. This layout is similar to the microbolometer design, and the thermal design considerations are the same.

The pyroelectric detector can be operated in short-circuit (current) mode or open-circuit (voltage) mode. The short-circuit current from the detector is²⁴

$$i = pA_s \frac{d(\Delta T)}{dt}, \quad (5.48)$$

where the current is proportional to the time derivative of the tempera-

ture change, $p = dP/dT$ is the pyroelectric coefficient in $[C/(m^2 \cdot K)]$, P is the polarization, A_s is the area of the detector in $[m^2]$, and (ΔT) is the temperature change in the material [Equation (5.39)]. The current-mode responsivity of the pyroelectric detector is then given by

$$\mathcal{R} = \frac{\epsilon p A_d \omega R_p}{G \sqrt{2(1 + \omega^2 \tau_{RC}^2)(1 + \omega^2 \tau_\theta^2)}}, \quad (5.49)$$

where ϵ is the surface absorption (emissivity), p is the pyroelectric coefficient, A_d is the area of the detector, R_p is the parallel resistance value of the loss (element) resistance and the pre-amplifier input resistance, G is the heat conduction coefficient, τ_{RC} is the electrical time constant, and τ_θ is the thermal response time.

Note that the device has a bandpass response, with a zero at the origin and two poles (electrical and thermal time constants). The term ‘zero’ is a differentiator (an ‘s’ operator in Laplace terminology), stemming from the differentiator in Equation (5.48) — this means that the detector has zero output at zero frequency. The first ‘pole’ at $1/\tau_\theta$ is the thermal time constant [Equation (5.39)], and the second pole at $1/\tau_{RC}$ results from the RC electrical time constant of the detector capacitance and the load resistor.

The pyroelectric detector has three noise sources: (1) Johnson noise, that is spectrally altered by the electronics RC time constant of the device (see Section 5.3.2), (2) temperature-fluctuation noise (see Section 5.3.6), and (3) read-out electronics interface circuit noise (see Section 5.3.7). These noise sources are uncorrelated and add in quadrature (see Section 5.3.12).

5.4.5 Thermoelectric detector

The thermoelectric detector senses the temperature change as a voltage change. The thermocouple is an example of the thermoelectric detector.³⁹ Two dissimilar metals or semiconductors, or a metal and a semiconductor, must be electrically and mechanically joined, such as by soldering. If the two junctions so formed are at different temperatures, a voltage will be generated according to the Seebeck effect. For small temperature differences the generated voltage is linearly proportional to the temperature difference. The Seebeck effect in semiconductor materials is more complex than in metals because the semiconductor behavior also influences the effect. The conceptual design of a thermoelectric detector is shown in Figure 5.5.

The thermoelectric detector has lower responsivity than the pyroelectric and bolometer detectors, but it does not require a bias current or optical signal chopping to form an image. Furthermore, the thermoelectric

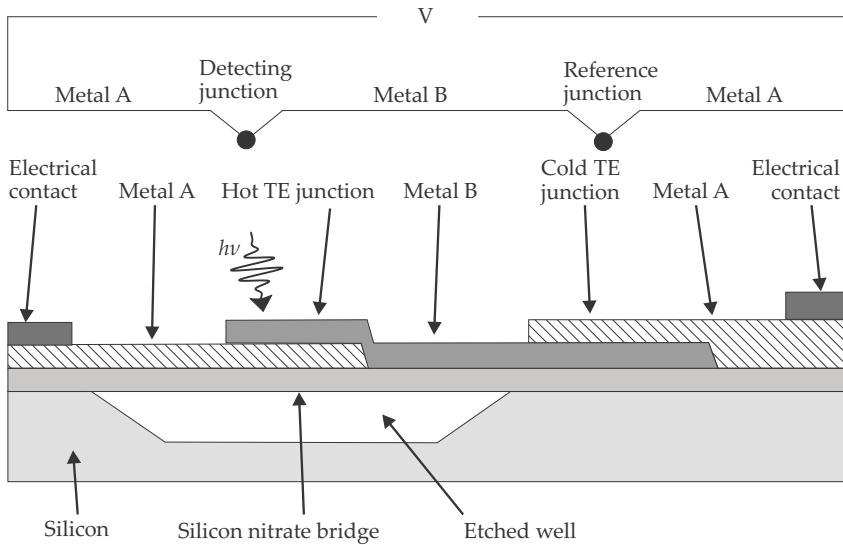


Figure 5.5 Thermoelectrical device layout (adapted²⁵).

detector has wider dynamic range than other detector techniques. In order to increase the output signal, more than one junction pair can be combined on one optical pixel, which is then called a thermopile.

When a metal rod is heated at the one end and cooled at the other end, the electrons at the hot end have more energy kT and thus greater velocity than at the cold end. There is a net diffusion of electrons from the hot end to the cold end, which leaves the positive ions in the hot region. This diffusion continues until the potential field so generated prevents the diffusion of more electrons.⁴⁰ A voltage is generated between the hot and cold ends of the rod. The ratio of the potential generated across the metal to a unit temperature difference is called the Seebeck coefficient $S = \Delta V / \Delta T$. By convention the sign of S represents the potential of the cold end relative to the hot end, but the cold end has a negative potential relative to the hot end. The Seebeck coefficient in a metal can be negative (Na, K, Al, Mg, Pb, Pd, Pt) or positive (Mo, Li, Cu, Ag, Au), depending on the diffusion of the electrons. The Seebeck coefficient in a p-type semiconductor is positive because in a p-type material the holes are the majority carriers. The Seebeck coefficient depends on the material temperature $S(T)$. It can be shown that the Seebeck coefficient for metals is given by (after some approximation)⁴⁰

$$S \approx -\frac{\pi^2 k^2 T}{3qE_F(0)}, \quad (5.50)$$

where k is Boltzmann's constant, T is the temperature in [K], q is the charge on an electron, and $E_F(0)$ is the Fermi energy at 0 K. The voltage across a

conductor with its end temperatures at T_0 and T_1 is determined by

$$\Delta V = \int_{T_0}^{T_1} S(T) dT. \quad (5.51)$$

Measuring the potential across a copper bar with electrical fly leads made of copper — and connected to the hot and cold end of the bar — will yield a zero voltage. The copper fly leads form the same potential difference, opposing the potential in the bar under test resulting in a zero voltmeter reading. If, however, the fly leads are made of another metal, the difference in two Seebeck potentials will be measured on the voltmeter:

$$\Delta V_{AB} = \int_{T_0}^{T_1} (S_A - S_B) dT. \quad (5.52)$$

Combining Equations (5.50) and (5.52) yields the metal thermocouple equation

$$V_{AB} = a\Delta T + b(\Delta T)^2, \quad (5.53)$$

where a and b are the thermocouple coefficients, and ΔT is the temperature with respect to the reference temperature of 273.16 K.

The responsivity of the thermoelectric detector is given by

$$\mathcal{R} = \frac{\epsilon N(S_1 - S_2)}{G \sqrt{1 + (\omega \tau_\theta)^2}}, \quad (5.54)$$

where ϵ is the surface absorption (emissivity), N is the number of junction pairs per pixel, S_1 and S_2 are the Seebeck coefficients of the two dissimilar materials, and G is the effective heat conduction coefficient [Equation (5.38)].

The thermoelectric detector has three noise sources: (1) Johnson noise (see Section 5.3.2), (2) temperature-fluctuation noise (see Section 5.3.6), and (3) read-out electronics interface circuit noise (see Section 5.3.7). However, the responsivity is relatively low, so the Johnson noise dominates.

5.4.6 Photon-noise-limited operation

Photon-noise-limited operation in thermal detectors occurs when the noise in the background and target photon flux is the limiting noise. This noise source is not in the detector or its electronics but is inherent in the signal. For a detector with unlimited spectral response (i.e., over all wavelengths) the photon-noise-limited D^* can be shown^{12,25} to be

$$D^* = \frac{\epsilon}{\sqrt{8\epsilon\sigma_e k(T_d^5 + T_e^5)}}, \quad (5.55)$$

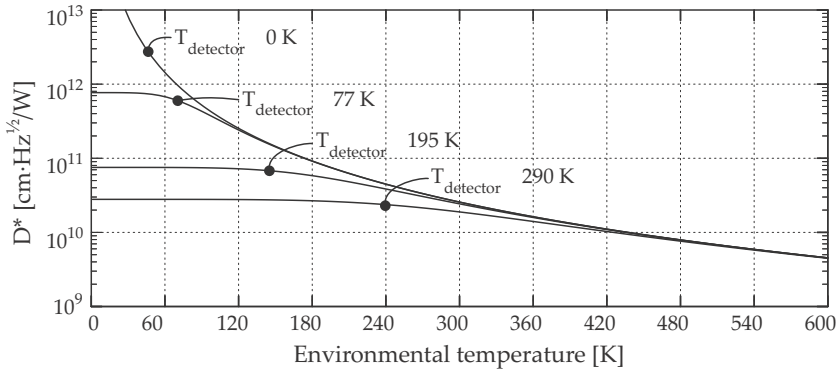


Figure 5.6 Photon-noise-limited D^* for thermal detectors with $\epsilon = 1$ and 2π sr field of view (adapted²⁵).

where ϵ is the surface absorption/emissivity, σ_e is the Stefan–Boltzmann constant, k is Boltzmann’s constant, T_d is the detector-element temperature, and T_e is the environmental temperature. The photon-noise-limited D^* is shown graphically in Figure 5.6.

Figure 5.6 indicates that a detector at room temperature, viewing a 280-K scene, can, at best, have a D^* value of $2 \times 10^{10} \text{ cm} \cdot \sqrt{\text{Hz}}/\text{W}$. If either the detector or the background is reduced to absolute zero, the D^* value is still no better than $2.8 \times 10^{10} \text{ cm} \cdot \sqrt{\text{Hz}}/\text{W}$. This derivation applies to a detector with a wide spectral band — if a narrow spectral band is used, the D^* will improve (see Section 9.7). These D^* values apply to a detector FOV of 2π sr. If a cold shield is used to limit the detector’s view of the ambient background, the D^* will improve by $1/\sin \theta$, where θ is the half-apex angle of the cold shield. In order for the cold shield to be effective, the shield must be considerably below room temperature.

Practical detectors may limit the spectral bandwidth with cold filters. The effect of the cold filter ($T_{\text{filter}} \ll T_e$) on D^* can be modeled by^{12,25}

$$D^* = \frac{\epsilon}{\sqrt{8\epsilon\sigma_e k(T_d^5 + F_{\lambda_1, \lambda_2})}}, \quad (5.56)$$

where

$$F_{\lambda_1, \lambda_2} = 2 \int_{\lambda_1}^{\lambda_2} \frac{h^2 c^3}{\lambda^6} \frac{e^x}{(e^x - 1)^2} d\lambda, \quad (5.57)$$

where $x = hc/(\lambda k T_e)$, T_e is the environment temperature, λ_1 is the filter short-cutoff wavelength, and λ_2 is the filter long-cutoff wavelength.

For a detector employed in a focal plane array the *photon-noise-limited*

NETD (noise limit imposed by the environment on the detector) is²⁵

$$NETD = \frac{8F_{\#}^2 \sqrt{2k\sigma_e \Delta f (T_d^5 + T_e^5)}}{\tau_s \sqrt{f_{\text{fill}} \epsilon A_d \frac{\partial M}{\partial T}}}, \quad (5.58)$$

where $F_{\#}$ is the detector cold shield f -number, Δf is the equivalent noise bandwidth, τ_s is the optics transmittance, f_{fill} is the focal plane array fill factor, ϵ is the spectrally flat device absorption (emissivity), $\frac{\partial M}{\partial T}$ is the temperature derivative of Planck's law, and A_d is the area of the detector.

5.4.7 Temperature-fluctuation-noise-limited operation

Temperature-fluctuation-noise-limited operation occurs when the device temperature variations limits noise performance (see Section 5.3.6). The temperature-fluctuation-noise-limited D^* of a thermal detector with wide spectral response can be shown^{12,25} to be

$$D^* = \sqrt{\frac{\epsilon^2 A_d}{4kT_d^2 G}}, \quad (5.59)$$

where ϵ is the spectrally flat device absorption (emissivity), A_d is the area of the detector, k is Boltzmann's constant, T_d is the device temperature, and G is the device thermal conductance between the sensing element and the environment.

If the detector is employed in a focal plane array, the *temperature-fluctuation-noise-limited* NETD (noise limit imposed by temperature fluctuations in the device) can be shown²⁵ to be

$$NETD = \frac{8F^2 T \sqrt{kG \Delta f}}{\tau_s f_{\text{fill}} A_d \epsilon \frac{\partial M}{\partial T}}, \quad (5.60)$$

where F is detector cold shield f -number, T is the device temperature, k is Boltzmann's constant, G is the device thermal conductance between the sensing element and the environment, Δf is the equivalent noise bandwidth, τ_s is the optics transmittance, f_{fill} is the focal plane array fill factor, A_d is the area of the detector, and ϵ is the device emissivity.

5.5 Properties of Crystalline Materials

To understand the photon-detection process, it is necessary to understand the quantum effects in the interaction between semiconductor structures and incident radiance fields. This section presents the basic theoretical aspects regarding semiconductor materials and their physical characteristics that allow the absorption of light.

5.5.1 Crystalline structure

Modern electronics technology employs crystalline semiconductor materials as a primary means to achieve its objectives. A crystal is a structured lattice arrangement of identical building blocks, where each block is an atom or a group of atoms. Crystalline structures are found in nature, but advances in crystal growth techniques allow production of near-perfect artificial crystalline structures designed to obtain specific properties and performance.¹ These crystals can be elemental, such as silicon or germanium, or they can be alloys, such as InSb or HgCdTe. Some alloys support composition ratios; for example, the $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ ternary alloy can be tuned to different bandgap values by setting the ratio of HgTe to CdTe.⁴¹ By setting the ratio, the detector's spectral range can be set to SWIR, MWIR, LWIR, and even beyond to 14–30 μm .

The structure of a crystalline material is defined by two important concepts, the lattice and the basis.¹ The lattice is a periodic geometrical construction of points in space where all of the points have exactly the same neighboring environment. The basis is a fundamental building block of atoms attached to each lattice point. This location of the atoms in the lattice structure provides the crystal with its unique physical, optical, and electronic properties.

Any lattice point R' can be obtained from any other lattice point R by the translation:

$$R' = R + m_1a_1 + m_2a_2 + m_3a_3, \quad (5.61)$$

where a_1 , a_2 , and a_3 are three primitive basis vectors, and m_1 , m_2 , and m_3 are integers. This construction is known as the Bravais lattice,⁴² and it allows the generation of any lattice by all possible combinations of the integers m_1 , m_2 , m_3 . The crystal structure remains invariant under translation through any vector that is the sum of integral multiples of the basis vectors.⁴³ There are 14 unique Bravais lattices in three-dimensional space. They are classified according to the relationship between the three primitive vectors and the angles (α , β , and γ) between them. A detailed discussion about crystalline structure can be found in Singh.¹

The periodicity in a lattice allows the definition of symmetry via a set of operations around a point: rotation, reflection, and inversion. The symmetry plays an important role in the electronic properties of the crystals — many physical properties of semiconductors are tied to the presence or absence of symmetry. For example, in the diamond structure (Si, Ge, C, etc.) inversion symmetry is present, whereas in the zinc blend structure (GaAs, AlAs, InAs, etc.) it is absent.¹

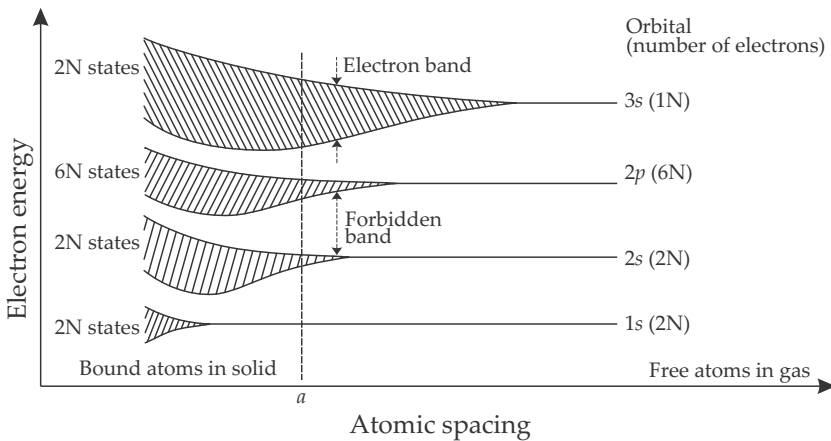


Figure 5.7 Band creation due to the decreasing distance between atoms for sodium (adapted^{45,46}).

5.5.2 Occupation of electrons in energy bands

The probability of finding an electron around an atom is given by the atomic orbitals (electronic state), a set of mathematical functions describing the behavior of one or a pair of electrons. The orbitals are designated by $1s$, $2s$, $2p$, etc., where each orbital represents a discrete, quantized energy level. Each orbital can contain at most two electrons, of opposite spin, by Pauli's exclusion principle. The atom's energy state is defined by its electronic distribution in the available orbital states. Orbitals represent increasing energy levels from the 'inner' to 'outer' orbitals ($1s$, to $2s$, to $2p$, to ...).⁴⁴

When atoms are bound in a crystal lattice, the energy levels of each atom are perturbed by the presence of other neighboring atoms, and by the lattice as a whole. Figure 5.7 shows the electronic energy states of the sodium atom as a function of interatomic spacing.⁴⁵ For large interatomic spacing (free atoms in gas), there is no perturbation from the neighboring atoms. As the spacing decreases, Pauli's exclusion principle prohibits electrons sharing the same energy orbitals. As a result, the orbital energy levels start to differ slightly between adjacent atoms — the fixed energy level for free atoms now becomes a range or 'band' of possible energy levels.⁴⁶ If a large number of atoms are involved, the splitting of energy levels will result in a quasi-continuous spread (a probability function), hence the notion of an energy band. Electrons can only occupy energy states within a band, never between the bands. The energy space between the bands is called the forbidden band or bandgap. This widening in the electronic energy probability distribution results from the solution of the orbital electronic wave function in the atomic-lattice context instead of the isolated-atom

context.⁴⁵ The width or spread of the band depends on the overlap of the individual atoms' energy orbitals — it affects the outer or higher energy orbitals (e.g., 3s) first before affecting the inner orbitals (1s).^{46,47} Figure 5.7 indicates four orbital energy bands at an atomic spacing of a . Note that for the 1s state the energy probability distribution is very narrow, whereas for the 3s band the probability distribution widens considerably.

The bandgap or atomic energy levels are commonly defined in units of electron volts [eV], the energy gained or lost by the charge of an electron moving across an electric field of one volt. The charge on one electron is q C. One volt is 1 J/C, hence one eV is quantity (q) J.

An important parameter in this theory is the Fermi level: electrons fill the orbital states from the lowest energy level up to a higher energy level E_f , called the Fermi Level. The probability of finding a *thermally excited* occupied energy state with an electron in an energy range E to $E + dE$ (under thermodynamic equilibrium) is given by the Fermi–Dirac distribution:

$$f(E) = \frac{1}{1 + e^{(E-E_F)/(kT)}}, \quad (5.62)$$

where T is the temperature in [K], E_F is the Fermi level in [J], and k is the Boltzmann constant. Note that kT is the thermal energy of an electron associated with the temperature T . The Fermi–Dirac distribution satisfies Pauli's exclusion principle and hence describes the distribution at all temperatures.⁴² At absolute zero temperature (0 K), all of the energy states are filled from the lowest state, with no vacant or unfilled states, up to the highest state defined by the Fermi level E_F . At nonzero temperatures, the filled states spread around the Fermi level, according to the Fermi–Dirac distribution. At higher temperatures, the spread is broader, as shown in Figure 5.8.

5.5.3 Electron density in energy bands

The classical view of an electron as a particle is not sufficient to describe the behavior of electrons in solid crystals. This analysis employs the wave model of the electron. In order to determine the number of electrons in an energy band, the number of allowable electron waves in the crystal must be determined. The number of electron waves in the crystal is determined by the number of energy states per unit energy, given by the density of states $D(E)$.^{10,45,46,48–50} The 'density of states' methodology developed here is also used in the derivation of thermal noise and Planck's law (see Sections 3.1 and 5.3.2).²³

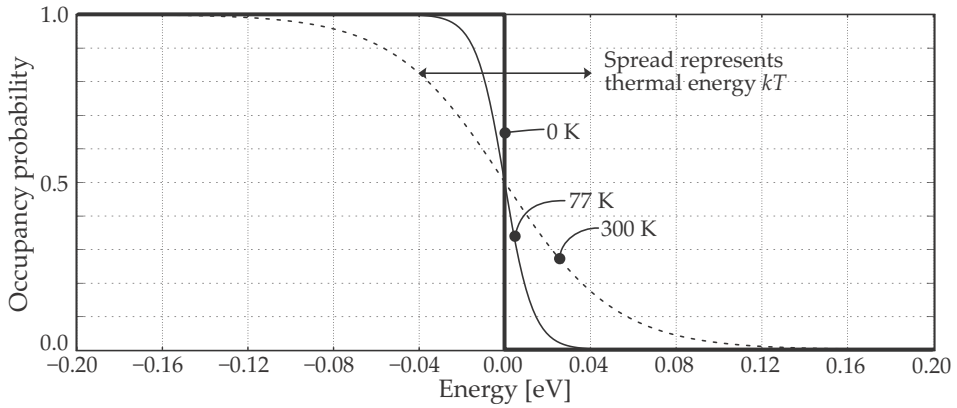


Figure 5.8 Fermi–Dirac energy probability distribution for thermally excited electrons, around the Fermi (50%) energy level.

The wave function for a free electron can be written as¹⁰

$$\psi(\mathbf{k}, \mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} = \cos(\mathbf{k}\mathbf{r}) + i \sin(\mathbf{k}\mathbf{r}), \quad (5.63)$$

where $\mathbf{k} = (\mathbf{k}_x + \mathbf{k}_y + \mathbf{k}_z)$ is the wave vector. The analysis is based on the principle that the electron wave must be one of the possible set of standing waves in the crystal — essentially a boundary constraint because the electron wave must have zero intensity at the crystal edges. Think of the electron wave vector \mathbf{k}_i as a resonant standing wave in the crystal, with twice the longest wavelength (lowest energy state) equal to the total length of the crystal L . This is a standing wave with nodes at each end of the crystal with a peak in the center of the crystal. The crystal size L corresponds to half the wavelength or π phase along the wave. Higher-order modes must be integer multiples of this half wave $\mathbf{k}_x = n_x\pi/L_x$ along the x direction, and likewise along y and z .

In the simplified model of a free electron, the energy for a free electron is given by the allowable standing-wave modes supported in a crystal cube with sides of length L .⁴⁸

$$\begin{aligned} E &= \frac{\hbar^2 \mathbf{k}^2}{2m^*} = \frac{\hbar^2}{2m^*} (\mathbf{k}_x^2 + \mathbf{k}_y^2 + \mathbf{k}_z^2) \\ &= \frac{h^2}{8m^* L^2} (n_x^2 + n_y^2 + n_z^2) = E_0 k_F, \end{aligned} \quad (5.64)$$

where \mathbf{k} is the wave vector with its three Cartesian coordinates, m^* is the effective electron mass, and $\hbar = h/(2\pi)$, where h is the Planck constant. This equation is an approximation⁴³ of the electron's energy near the band edges E_C and E_V , acceptable within the primary concern of this analysis. The value $E_0 = h^2/(8m^* L^2)$ is the lowest energy state in the crystal (longest

wavelength). $k_F^2 = (n_x^2 + n_y^2 + n_z^2)$, where n_i are the indices of the reciprocal lattice points (1/distance) inside the sphere with radius k_F . This sphere contains all of the modes associated with indices less than n_i — that is, all of the standing waves with longer wavelengths (lower energy). The number of modes that can be sustained in a sphere with radius k_F is given by all of the modes starting with the shortest wavelength, then including all of the modes with longer wavelengths up to wavelength k_F .

The number of electrons that can be accommodated in states with energy E or less is⁴⁸

$$N(E) = 2 \left(\frac{1}{8} \right) \left(\frac{4\pi k_F^3}{3} \right) = \frac{\pi}{3} \left(\frac{E}{E_0} \right)^{3/2}, \quad (5.65)$$

where the factor of 2 allows for the positive and negative electron spin for each state, and $4\pi k_F^3/3$ is the volume of the sphere. The factor of 8 is present because only the positive integers are considered. It can then be shown^{45,48} that the density of electrons *with energy less than E* per unit volume is

$$D(E) = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2} = \frac{\pi}{2} \left(\frac{8m^*}{\hbar^2} \right)^{3/2} E^{1/2}. \quad (5.66)$$

Having N electrons in the energy band per unit of volume, the condition determines that E_F is given by⁴⁹

$$\int_0^{E_F} D(E) dE = N. \quad (5.67)$$

The Fermi level is obtained by solving the integral:

$$E_F = \left(\frac{\hbar^2}{2m^*} \right) (3\pi^2 N)^{2/3} = \left(\frac{\hbar^2}{8m^*} \right) \left(\frac{3N}{\pi} \right)^{2/3}. \quad (5.68)$$

The band that is normally *fully filled* with electrons at 0 K is called the valence band, whereas the upper unfilled band is called the conduction band.

From Equation (5.66) the density of electrons per unit volume in the conduction and valence bands becomes¹⁰

$$D_C(E) = \frac{\pi}{2} \left(\frac{8m^*}{\hbar^2} \right)^{3/2} (E - E_C)^{1/2} \text{ for } (E > E_C) \quad (5.69)$$

$$D_V(E) = \frac{\pi}{2} \left(\frac{8m^*}{\hbar^2} \right)^{3/2} (E_V - E)^{1/2} \text{ for } (E < E_V). \quad (5.70)$$

Figure 5.9(a) shows the relationship between electron energy as a function of wavenumber along one dimension. The figure shows a wave vector

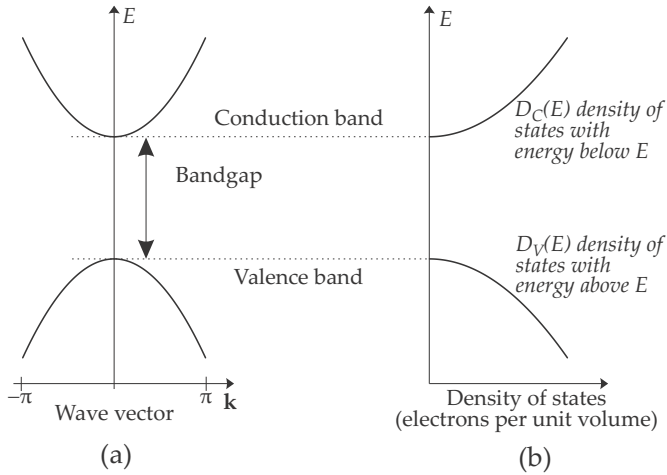


Figure 5.9 Crystalline material: (a) relationship between electron energy and wave number k_i , and (b) density of states for a free electron in a semiconductor.

for a single atom in the crystal. These wave vectors repeat exactly along the lattice positions, a requirement for stable electronic wave solutions. Figure 5.9(b) shows the density of electron states in the conduction and valence bands of a semiconductor. Both of these functions are parabolic in shape.

5.5.4 Semiconductor band structure

Electron dynamics in semiconductors is modeled by the Schrödinger equation, the solution of which describes the band structure of the material. Given the large number of atoms in a material, the electrons are subjected to a very complex potential profile. Atoms vibrate, adding a temporal variation to the atomic potential profile. Finally, electrons interact with each other. Solving the Schrödinger equation is simplified in crystalline materials because the fixed spatial structure in the crystal presents a periodic potential background, satisfying the Bloch theorem.^{10,51}

The Schrödinger equation is given by¹

$$\left[\frac{-\hbar^2}{2m_0} \nabla^2 + U(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}), \quad (5.71)$$

where m_0 is the mass of the electron, $U(\mathbf{r})$ is the potential seen by the electron, $\psi(\mathbf{r})$ is the wave function describing the electronic state, and E is energy. The first term represents the free electron's wave function. The second term represents the dependency on all of the other electrons and atoms in the material. For crystalline materials, the potential $U(\mathbf{r})$ has

a lattice periodicity R , hence $U(\mathbf{r}) = U(\mathbf{r} + R)$. Because of the periodicity shown by $U(r)$, the wave function will also be periodic⁵¹ $|\psi(\mathbf{r})|^2 = |\psi(\mathbf{r} + R)|^2$. The physical meaning of ψ is that $|\psi|^2 dx dy dz$ represents the probability of finding an electron in the volume element $dx dy dz$ in the vicinity of position (x, y, z) .⁵² It can be shown¹ that the wave function is spread across the entire sample and has equal probability ($\psi^* \psi$) at every point in the lattice. Note that $\psi(\mathbf{r})$ is a probability, hence the probability is periodic over the lattice.¹

The electronic wave function is given by the Bloch functions^{1,10}

$$\psi(\mathbf{k}, \mathbf{r}) = u_n(\mathbf{k}, \mathbf{r}) e^{i\mathbf{k}\mathbf{r}}, \quad (5.72)$$

a lattice-periodic set of plane waves with a space-dependent amplitude factor $u_n(\mathbf{k}, \mathbf{r})$. This important result shows that electrons in the lattice can be described by means of a wave function vector \mathbf{k} and energy $E(\mathbf{k})$, as shown in Figure 5.9(a).

5.5.5 Conductors, semiconductors, and insulators

Materials are classified as electrical insulators, semiconductors, or conductors, depending on the width of the forbidden energy gap between two outer energy bands (see Figure 5.10). Insulators have a large bandgap (6 eV in the case of diamond) that prevents thermionic excitation at environmental temperatures, and consequently no electrons are present in the conduction band. The valence band and conduction band overlap in conductors, allowing valence electrons to move freely through the material, thus, resulting in electrical conductivity. Semiconductor materials can act as insulators or conductors, depending on the temperature of the material. At low temperature, the thermally excited electrons [see Equation (5.62) and Figure 5.8] do not have sufficient energy to be excited to the conduction band (insulator). At higher temperatures, the thermal excitation may excite a large number of electrons to the conduction band (conductor). Examples of semiconductor materials are Si, Ge, GaAs, InSb, and HgCdTe, all of which are commonly used IR-detector materials.

If a crystalline material is tetravalent (each atom has four neighbors as in silicon), the valence band is filled and occupied by four electrons. These electrons are closely bound to the atom (in its physical position in the lattice) and do not contribute to conduction. At low temperatures, the higher energy conduction band is empty. If the temperature is raised and the electrons in the valence band gain sufficient energy to cross the forbidden energy gap, these electrons are excited to the conduction band. The electrons in the conduction band are not closely bound to the atom; they belong to the crystal as a whole. These electrons can therefore move

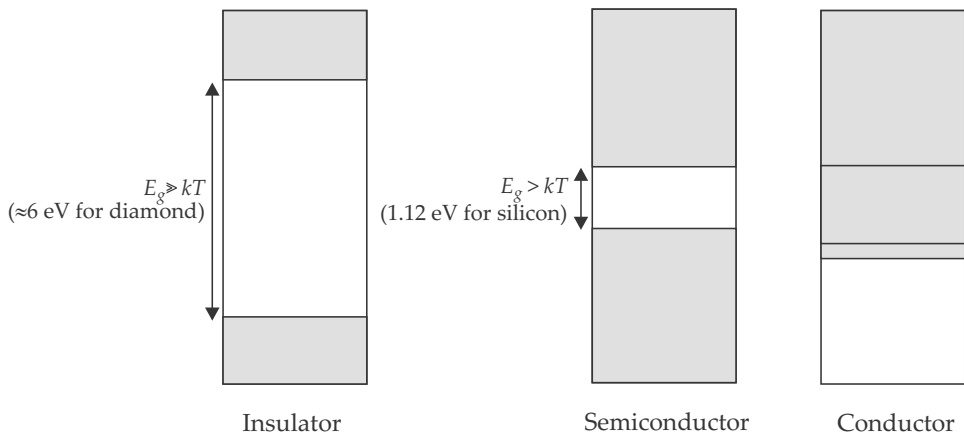


Figure 5.10 Energy bands for isolators, semiconductors, and conductors.

about freely and contribute to the conduction of electric current. Both electrons e^- and holes h^+ (the absence of an electron — an empty energy level) contribute to conduction in a semiconductor.

5.5.6 Intrinsic and extrinsic semiconductor materials

In *intrinsic* (pure) semiconductor materials there are no impurities, and for *every atom* in the lattice the number of electrons in the conduction band exactly matches the number of holes in the valence band. If the semiconductor is doped with a small and controlled amount of a foreign atomic element (impurities), the material is called an *extrinsic* semiconductor. The dopant atoms replace the host atoms in the crystal lattice with considerable effect on the surrounding atoms' bonds and electronic behavior, and therefore the material properties. The impurities added to tetravalent crystal semiconductors (e.g., silicon) can have five valence electrons (donors) or three valence electrons (acceptors). When these atoms are located in the tetravalent crystal lattice, the number of electrons and holes in the valence band do not match the surrounding tetravalent atoms.

Pentavalent donor impurities (phosphorus) add additional electron-hole pairs to the crystal. Four of the phosphorus atom's electrons are tied up in the covalent bonds to neighboring silicon atoms. The fifth electron is free to move in the conduction band, whereas the fifth hole is bound to the pentavalent atom's location in the lattice [see Figure 5.11(b)]. The donor doped material is called *n-type* material. In n-type material, the mobile electrons are the majority carriers, and the bound holes are the minority carriers.

Trivalent acceptor impurities (e.g., boron) remove electron-hole pairs

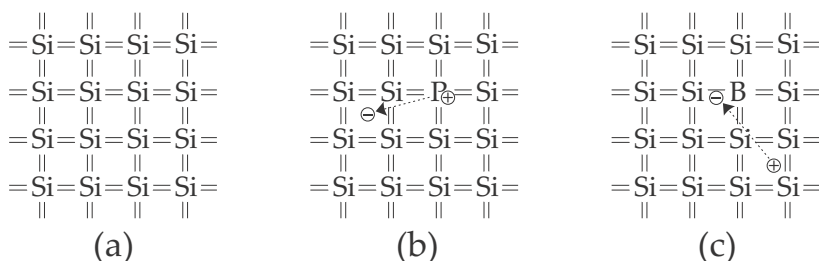


Figure 5.11 Silicon semiconductor lattice at 0 K: (a) intrinsic, (b) extrinsic with donor doping (n-type), and (c) extrinsic with acceptor doping (p-type).^{49,50}

from the crystal. The covalent bonds with the neighboring silicon atoms require four electrons, whereas the trivalent atom can only provide three. A valence-band electron becomes ‘caught’ in this fourth covalent bond — it is not available for excitation to the conduction band. The silicon atom providing this trapped electron now has one less electron (a hole). The hole (absence of an electron) in the valence band can freely move through the material because an electron fills the hole but in the process it creates a new hole in another atom [see Figure 5.11(c)]. The acceptor-doped material is called *p-type* material. In p-type material, the mobile holes are the majority carriers, and the bound electrons are the minority carriers.

Only a small amount of energy (0.01–0.05 eV) is required to elevate the ‘extra’ electron from the donor’s energy level E_D into the conduction band [see Figure 5.12(b)]. Similarly, a small amount of energy will excite an electron from the valence band into the acceptor level E_A (the missing electron in the covalent bond) [see Figure 5.12(c)]. The impurity atoms therefore introduce ‘allowable’ energy states in the otherwise forbidden bandgap.

Donor doping introduces electrons in the conduction band but captures the holes in fixed lattice locations. Acceptor doping introduces holes in the valence band but captures the electrons in fixed lattice locations. The free electrons and holes generated by doping do not have their free counterparts available for conduction, as in a metal where both free electrons and holes are present. Electron-hole pairs formed by thermal excitation support conduction by free electrons in the conduction band and by free holes in the valence band.

Figure 5.12 shows electron excitation by photon absorption, but note that thermal excitation has the same effect in the conduction band. Electron excitation by thermal means or by photon means are indiscernible once the electron is excited. Thermally excited electrons normally interfere with the device’s operation — they must be minimized. For this reason,

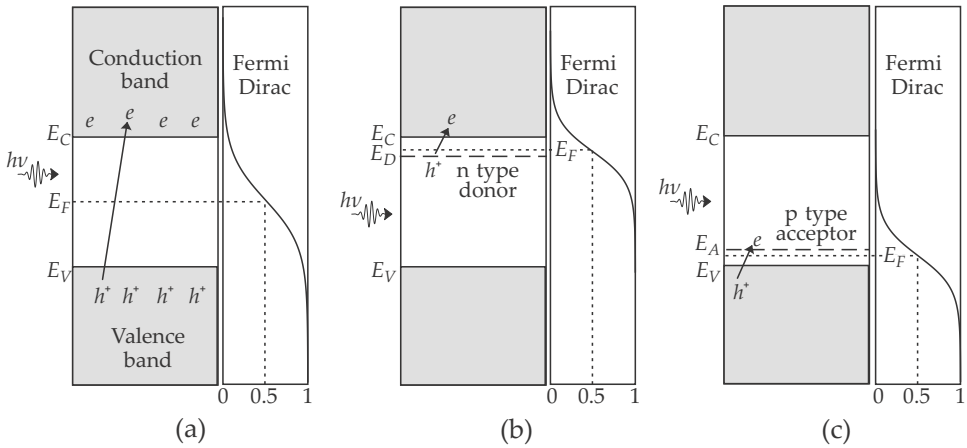


Figure 5.12 Energy bands and Fermi–Dirac distributions for: (a) intrinsic semiconductor, (b) n-type extrinsic semiconductor, and (c) p-type extrinsic semiconductor.

semiconductor detectors with small bandgaps are cooled down to reduce thermal excitation.

It can be shown that for intrinsic material, the Fermi energy level is more or less halfway between the valence and the conduction bands:⁴⁸

$$E_F = \frac{E_g}{2} + \frac{3kT}{4} \log_e \left(\frac{m_h^*}{m_e^*} \right), \quad (5.73)$$

where E_g is the bandgap, k is the Boltzmann constant, T is the temperature, m_h^* is the effective hole mass, and m_e^* is the effective electron mass. The E_F deviation from the exact center depends on the temperature and the values of the effective hole and electron masses. The Fermi level for extrinsic materials is shifted, depending on the nature of the doping^{49,50} (see Figure 5.12).

In intrinsic detectors, the donor concentration n_d and acceptor concentration n_a are relatively low compared to intrinsic concentrations n_i . Whereas these impurity levels are too small to have much significance on the optical properties of the device, they do affect the electronic device characteristics of the detector. Examples of intrinsic detectors are silicon (Si), germanium (Ge), indium-antimonide (InSb), and mercury-cadmium-telluride (HgCdTe). The detector properties of the intrinsic detector is determined by the energy bandgap between the conduction band and the valence band of the material (Figure 5.12).

Extrinsic detectors employ donor and acceptor concentrations of 10^2 to 10^4 times *higher* than normally used for semiconductor device operation. Such high concentrations are required to achieve a reasonable responsivity performance because only the donor sites or acceptor sites can

excite electrons to higher energy bands — compared to intrinsic detectors, where every atom can excite an electron. At these high doping levels, the material can only be used in photoconductive applications (Section 5.8). Examples of extrinsic detectors are Si:Hg, Si:Ga, Ge:Cu, or Ge:Zn, where the first symbol indicates the detector bulk material, and the second symbol indicates the doping element. The detection cutoff wavelength λ_c of an extrinsic detector is determined by the energy difference between the dopant energy level and the appropriate band (conduction or valence) of the bulk material; see Figure 5.12.

At thermal equilibrium, the intrinsic material hole concentration p (in the valence band) equals the electron concentration n (in the conduction band), which is also the intrinsic carrier concentration n_i . The (temperature-dependent) intrinsic carrier concentration is given by^{2,10,16,50}

$$n_i^2 = np = 4 \left(\frac{2\pi kT}{h^2} \right)^3 \exp \left[\frac{-E_g}{kT} \right] (m_h^* m_e^*)^{3/2}. \quad (5.74)$$

For doped material the concentrations across the p-n junction are

$$n_i^2 = n_n p_n = n_p p_p, \quad (5.75)$$

where p_n is the hole concentration in n-type material in $[\text{cm}^{-3}]$, p_p is the hole concentration in p-type material in $[\text{cm}^{-3}]$, n_n is the electron concentration in n-type material in $[\text{cm}^{-3}]$, and n_p is the electron concentration in p-type material in $[\text{cm}^{-3}]$. The dopant impurity sites will normally be ionized — holes left behind in the valence band for electrons excited to the acceptor sites, or from electrons excited from the donor sites to the conduction band — thus supporting the free movement of electrons and holes in the material. Furthermore, thermal excitation is also normally low (few thermally excited electrons in the conduction band). Under these conditions the electron concentration in an n-type material n_n is equal to the donor concentration n_d because the donor electrons are the only electrons in the conduction band. It then follows that

$$p_n = n_i^2 / n_n = n_i^2 / n_d. \quad (5.76)$$

Likewise, the hole concentration in p-type material p_p is equal to the acceptor concentration n_a , hence

$$n_p = n_i^2 / p_p = n_i^2 / n_a. \quad (5.77)$$

5.5.7 Photon-electron interactions

When a photon is absorbed, two types of electron transitions can occur: intraband and interband transitions,⁵³ as shown in Figure 5.13 (the detailed

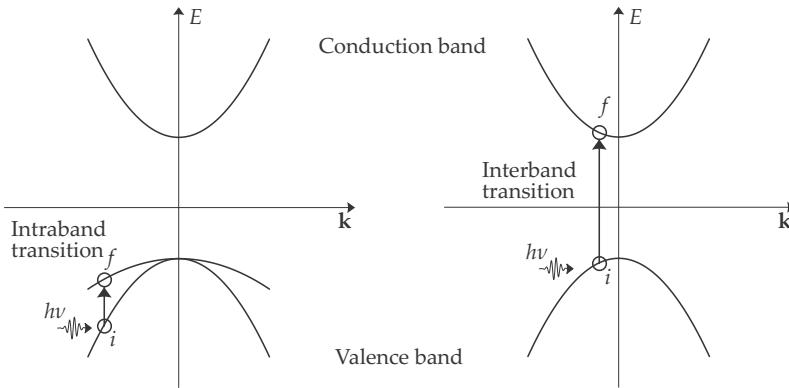


Figure 5.13 Intraband and interband transitions of an electron from an initial state i to a final state f .

situation is much more complex than this). Intraband transition happens when the electron jumps from one state to another in the same band. This phenomenon is common in quantum well devices but is outside the scope of this book.

Optically induced interband transitions occur when the absorption of a photon excites an electron across a forbidden energy gap to a higher energy state. Interband transitions are present in all solids.⁵³ In a semiconductor material, the interband transition excites an electron from an occupied state in the valence band (leaving behind a hole) into an unoccupied excited state in the conduction band. Note the following requirements and constraints regarding interband transitions:^{54,55}

1. The photon energy must exceed the material energy bandgap $E_{transition} = h\nu \geq E_g$.
2. The transitions can be either direct or indirect. A direct transition occurs when the crystal momentum is conserved such that $\mathbf{k}_i = \mathbf{k}_f$, where \mathbf{k}_i and \mathbf{k}_f are the initial and final wave vectors, respectively. An indirect transition involves a phonon, and momentum is not conserved. In this case the initial and final wave vectors differ by a value equal to the phonon wave vector \mathbf{q} , such that $\mathbf{k}_i = \mathbf{k}_f \pm \mathbf{q}_{phonon}$. The transitions are shown in Figure 5.14.

A *phonon* is a quantum of energy associated with an electron wave's interaction with a crystal lattice — it is essentially vibrational heat in the crystal lattice.^{1,56}

3. Certain transitions are forbidden, as regulated by selection rules.

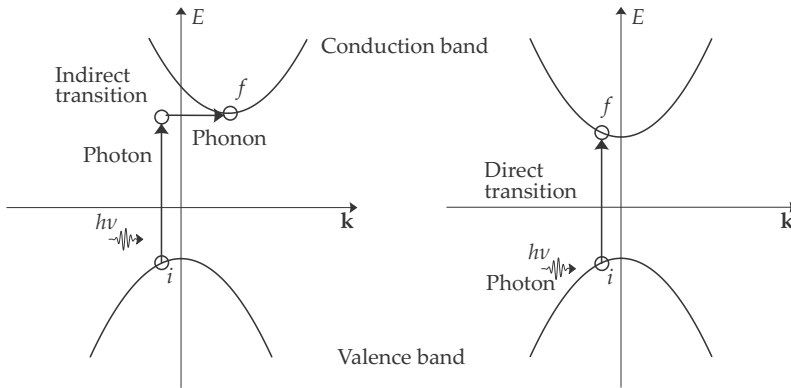


Figure 5.14 Interband transitions: (a) indirect transition involving a phonon and a photon, and (b) a direct transition.

4. Subject to the Pauli exclusion principle, an interband transition occurs from an occupied state below the Fermi level to an unoccupied state above the Fermi level.
5. If the energy gap $E(k)$ between two bands is near-constant over a wide range in k , photon-induced interband transitions occur more effectively. It means that there are many initial and final states resonant with this photon energy.

In photon detectors the transition near the smallest energy gap between the valence and conduction bands is of main interest. This occurs where $k = 0$, which is designated as the Γ point.

5.5.8 Light absorption in semiconductors

The solution of the wave equation through a dielectric medium yields a complex index of refraction $n = n_r + i n_i$, of which both terms alter the wave's properties^{10,43,49,55,57} (see also Sections 3.4.3 and 4.2.1). The real refractive index component changes the wave's velocity (and hence the wavelength), leading to Snell's law (the phase velocity across the wavefront varies, changing the wavefront's propagation direction). The imaginary component changes the wave's field strength (radiance). An increase in the number of interacting particles in the medium (electrons in the case of metals and semiconductors) leads to an increase in the imaginary refractive index. The imaginary refractive index component is related to the absorption coefficient of the medium¹⁰ by $\alpha\lambda = 4\pi n_i$. Photon absorption is a very complex physical process; the description given here provides only a brief insight into the underlying processes.

Light absorption occurs when an incident photon interacts with an electron in the valence band. The electron absorbs the photon's energy and is excited to the conduction band, leaving a hole behind. The radiance of a light beam propagating along the x direction in a material is given by $L_v(x) = L_v(0)e^{-\alpha_v x}$, where α_v is the material spectral absorption coefficient [see Equation (4.4) in Section 4.2.1]. The balance of the photons $L_v(0) - L_v(x)$ are absorbed in the material and converted to electrons.

The spectral absorption coefficient has different modes for the various different excitation transition processes,^{10,43,52,55} resulting in the characteristic shapes in the spectral absorption coefficient, Figure 5.15. Direct transition materials (e.g., GaAs, InP, InSb, and HgCdTe) have absorption coefficients of the form^{2,56}

$$\alpha_v \propto \sqrt{h\nu - E_g} \approx \alpha_{\lambda_c} + \alpha_0 \sqrt{h\nu - E_g} \quad \text{for } h\nu > E_g. \quad (5.78)$$

The energy structures in indirect semiconductors (e.g., Si and Ge) are laterally displaced, hence phonons are also involved in the excitation process. Indirect semiconductors have absorption coefficients of the form^{2,10,43,56}

$$\alpha_v \propto (h\nu - E_g \pm E_p)^2 \approx \alpha_{\lambda_c} + \alpha_0 (h\nu - E_g \pm E_p)^2 \quad \text{for } h\nu > E_g. \quad (5.79)$$

The free-carrier absorption coefficient is proportional to the electron and hole carrier concentration because each carrier site acts as an absorption site. At energy levels below the bandgap, an exponential absorption takes place, known as the Urbach tail,^{7,10,48,52,56,58-60} where

$$\alpha_v = \alpha_{\lambda_c} \exp \left[\frac{h\nu - E_g}{kT} \right] \quad \text{for } h\nu < E_g. \quad (5.80)$$

The spectral shapes of these absorption processes are shown in Figure 5.16.

The typical absorption coefficient curves shown in Figure 5.15 reflect specific material samples under specific conditions; different samples and measurements may differ. The absorption coefficient is a function of temperature, crystal orientation, and impurity concentration (intrinsic or extrinsic material).⁶¹

From Equation (4.4) it follows that the flux at depth d into the detector will be $\Phi(d) = \Phi(0)(1 - \rho)e^{-\alpha d}$. Most of the photons are absorbed within a distance of $1/\alpha$ into the absorbing material. For large absorption coefficients, this occurs within a thin layer on the skin of the detector. From the absorption graphs in Figure 5.15, it follows that photons with longer wavelength are absorbed deeper into the material, whereas photons with shorter wavelength are absorbed nearer the surface of the material.

What happens to the photons with wavelengths exceeding λ_c that are not absorbed in the material? The material normally has low attenuation

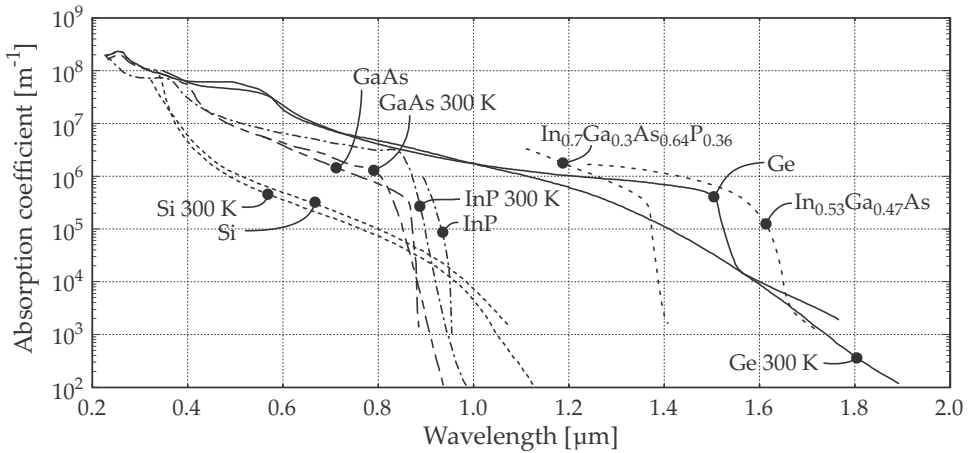


Figure 5.15 Photon absorption coefficient for intrinsic materials (adapted^{1,10,62}).

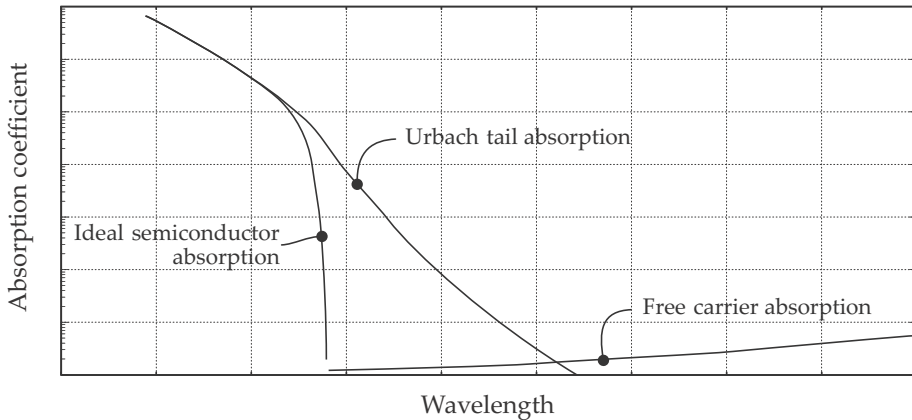


Figure 5.16 Photon absorption coefficient: ideal, Urbach tail, and free-carrier components. (adapted⁵⁶).

at wavelengths beyond λ_c . This fact is often employed to construct ‘sandwich’ detectors that are located in close proximity behind one another. Each detector in the sandwich is made of a different material, sensitive in a different spectral band. The detectors are electrically isolated from each other; i.e., independent detectors.

Every atom in an intrinsic material can absorb photons, but only the impurity locations in an extrinsic material can absorb photons. The absorption coefficients in intrinsic detector materials are higher than in extrinsic detector materials because there are more absorption sites in intrinsic materials than in extrinsic materials.

5.5.9 Physical parameters for important semiconductors

Tables A.5 and A.6 contain a summary of physical parameters for selected intrinsic and compound semiconductors. Material properties are well documented.⁶³ It is important to note that all parameters are to be used in calculations regarding direct transitions near the Γ point, at the minimum bandgap.

5.6 Overview of the Photon Detection Process

5.6.1 Photon detector operation

If the energy of the photon exceeds the energy bandgap of the material, an absorbed photon creates an electron-hole pair in the semiconductor material. The excited electron-hole pair recombines after some time, the average value of which is the carrier lifetime. The carrier lifetime is not a material property but rather a function of the application. Detectors that employ this phenomenon are called photon detectors.

Photoconductive detectors operate on the principle that photons are absorbed in the bulk of the detector material, and an electron-hole pair is formed. The electron-hole pair separates and contributes to electrical current conduction, thereby lowering the bulk resistance of the detector. The photon-induced conductance change in resistance can be detected by an external electronic circuit.

Photovoltaic detectors operate on the principle that electron-hole pairs formed by photon absorption in the depletion region of a p-n diode are accelerated across the depletion region. This electron-hole pair contributes to current flow by injecting minority carriers, thereby causing current flow through the depletion region. The photocurrent flows through the depletion region, under the built-in bias that exists in the depletion region. The photocurrent flow can be sensed by an external electronic circuit.

5.6.2 Carriers and current flow in semiconductor material

Current flow in semiconductors may be due to diffusion flow or drift flow by the carriers (holes and electrons). *Diffusion* movement of carriers results from the equalization of nonuniform densities of charges. The principle is the same as the diffusion of gas particles to remove nonuniformity in a gas distribution. Whenever a gradient exists in charge densities, the diffusion current will tend to remove such gradients. Diffusion current does not require an external field (electric voltage or gas pressure). The diffusion

current density is given by^{16,48}

$$J = -qD \frac{dn_c}{dx}, \quad (5.81)$$

where J is the diffusion current density in $[A/m^2]$, q is the electronic charge in $[C]$, D is the diffusion constant in $[m^2/s]$, and dn_c/dx is the carrier density gradient in the material.

Drift movement of carriers occurs when charged carriers are accelerated under an electric field by Ohm's law:

$$J_d = \sigma \mathcal{E} = nq\mu \mathcal{E}, \quad (5.82)$$

where J_d is the drift current density in $[A/m^2]$, σ is the material electrical conductivity in $[\Omega^{-1}m]$, \mathcal{E} is the electric field strength in $[V/m]$, nq is the charge density in $[C/m^3]$, and μ is the charge mobility in the material in $[m^2/(V \cdot s)]$.

The diffusion constant D and the charge mobility μ are statistical thermodynamic phenomena related by the Einstein equation

$$\frac{D_e}{\mu_e} = \frac{D_h}{\mu_h} = \frac{kT}{q}. \quad (5.83)$$

In the case where a potential gradient and a concentration gradient exist at the same time in the material, the total electron current is then

$$J_e = n_e q \mu_e \mathcal{E} + q D_e \frac{de}{dx}, \quad (5.84)$$

where de/dx is the electron density gradient. If dh/dx is the hole density gradient, the total hole current is

$$J_h = n_h q \mu_h \mathcal{E} + q D_h \frac{dh}{dx}. \quad (5.85)$$

The first term in these equations is the drift current through the material and the second term is the diffusion current. The drift and diffusion currents are caused by different forces and flow independently from each other.

5.6.3 Photon absorption and majority/minority carriers

When a material is illuminated with a constant photon flux Φ_q for a time period t , then the absolute change in number of holes equals the absolute change in number of electrons:

$$\Delta e = \Delta h = \Phi_q t. \quad (5.86)$$

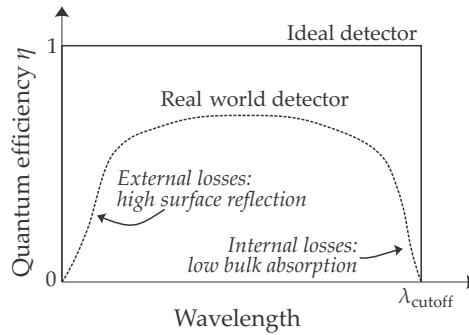


Figure 5.17 Comparison between ideal and actual photon detector quantum efficiency (adapted²).

If this absorption took place in an n-type material, the *percentage* change in the number of electrons (majority carriers) is very small ($\Delta e \ll n$) but the percentage change in the number of holes (minority carriers) is very large ($\Delta h > p$). For practical purposes, the percentage change in majority carriers is so small that it can be ignored. The percentage change in minority carriers is significant and can effectively be regarded as the primary signal. Hence, it is generally accepted that the signal is formed by the optical injection of *minority carriers*.

5.6.4 Quantum efficiency

The relationship between the number of incident photons on the detector and the number of generated electrons is called the quantum efficiency η . Quantum efficiency describes how effectively a photon detector can convert the incoming flux to current. Quantum efficiency has a value between zero and one, often expressed as a percentage.

Quantum efficiency has two components: (1) internal quantum efficiency indicating how well the detector material can convert photons to electrons, and (2) external quantum efficiency indicating the fraction of photons incident on the external surface of the detector that penetrate the material and are not reflected off the surface.

A detector's quantum efficiency is not constant but varies with wavelength, as shown in Figure 5.17. These deviations are particularly apparent for wavelengths around λ_c and for $\lambda \ll \lambda_c$. At longer wavelengths, the quantum efficiency is limited by the material's absorption coefficient (internal quantum efficiency). At shorter wavelengths, the quantum efficiency is determined by the reflection losses from the surface of the detector (external efficiency).

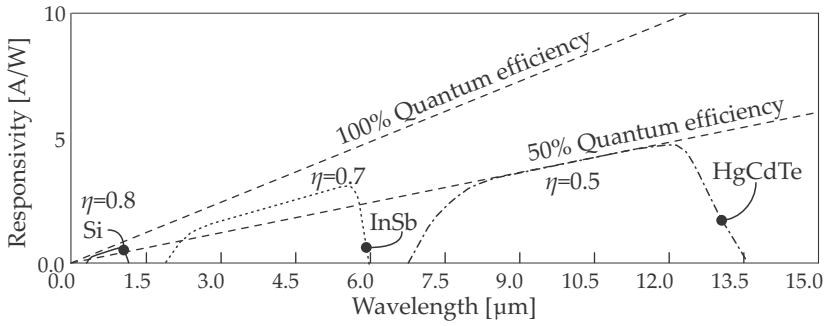


Figure 5.18 Typical detector spectral responsivity and quantum efficiency.

The material surface reflects a portion ρ of the total incoming flux by Fresnel reflection (see Section 3.4.4), affecting the external quantum efficiency. The value of ρ depends on the index of refraction of the detector material and is given by

$$\rho = \left(\frac{n_2 - n_1}{n_2 + n_1} \right)^2, \quad (5.87)$$

where n_2 is the refractive index of the detector material, and n_1 is the refractive index of the surrounding medium (normally air). For silicon and germanium detectors the reflection can be as high as 40%. Special anti-reflection coatings are normally applied to the surface to reduce the external reflection.

The quantum efficiency — the ratio of charge carriers generated to the incoming photon count — can be calculated from²

$$\frac{J}{qE_q} = \eta = (1 - \rho)(1 - e^{-\alpha d}), \quad (5.88)$$

where J is the current density in $[A/m^2]$, q is the electronic charge, E_q is the photon irradiance in $[q/(s \cdot m^2)]$, ρ is the semiconductor Fresnel reflectance, α is the detector material absorption coefficient (see Section 5.5.8), and d in $[m]$ is the depth over which absorption takes place in the material (i.e., depth of the active region in the detector). It is evident that the quantum efficiency increases when the reflectance decreases or where the absorption thickness αl_x is high (see Section 4.2.4). The material absorption coefficient depends on the wavelength, as shown in Figure 5.15.

The peak quantum efficiency in high-performance detectors often exceeds 0.8. A high quantum efficiency is normally obtained by an optimized anti-reflection coating on the detector surface. Figure 5.18 shows typical spectral responsivity curves for three photon detector types. Also shown are the quantum efficiencies for each detector as well as the 100% quantum

efficiency line [Equation (5.5)]. Note that the photon detector responsivity increases toward longer wavelengths, as shown in Figure 5.1. This stems from the fact that the photon energy decreases toward longer wavelengths and that larger numbers of photons (hence the larger detector current) are required to provide one watt of optical power.

The quantum efficiency for extrinsic detectors is lower than the quantum efficiency of intrinsic detectors because there are many more possible excitation locations in intrinsic detector materials than in extrinsic detector materials. Typical absorption coefficients for extrinsic materials are less than 10^3 m^{-1} , compared to $10^3\text{--}10^7 \text{ m}^{-1}$ for intrinsic materials.

Figure 5.19 shows a comparison of various detectors' spectral D^* performance. The caption in the original publication¹² is as follows:

Comparison of the D^* of various commercially available IR detectors when operated at the indicated temperature. Chopping frequency is 1000 Hz for all detectors except the thermopile (10 Hz), thermocouple (10 Hz), thermistor bolometer (10 Hz), Golay cell (10 Hz), and pyroelectric detector (10 Hz). Each detector is assumed to view a hemispherical surrounding at a temperature of 300 K. Theoretical curves for the background-limited D^* (dashed lines) for ideal photovoltaic and photoconductive detectors and thermal detectors are also shown. PC, photoconductive; PV, photovoltaic; PE, photoemissive; and PEM, photoelectromagnetic detector.

Note that some detectors have quite high D^* (high quantum efficiency) values compared to the theoretical limit.

5.7 Detector Cooling

Thermally excited carriers have energy kT , which may exceed the bandgap in a detector [Equation (5.62)]. The thermally excited carriers contribute to detector noise and reduce the performance of the detector. The objective with detector cooling is to achieve $kT \ll E_g$. It is evident that materials with wide bandgaps are less prone to thermal carrier excitation. Long-wavelength detectors generally operate at temperatures such as 77 K or even down to 4 K. Modern detector development, and particularly the use of HgCdTe, enabled LWIR detectors to operate at somewhat higher temperatures (80 K) than older extrinsic detectors (4 K). Current research is pursuing detector technologies that would enable even higher operating temperatures.⁷

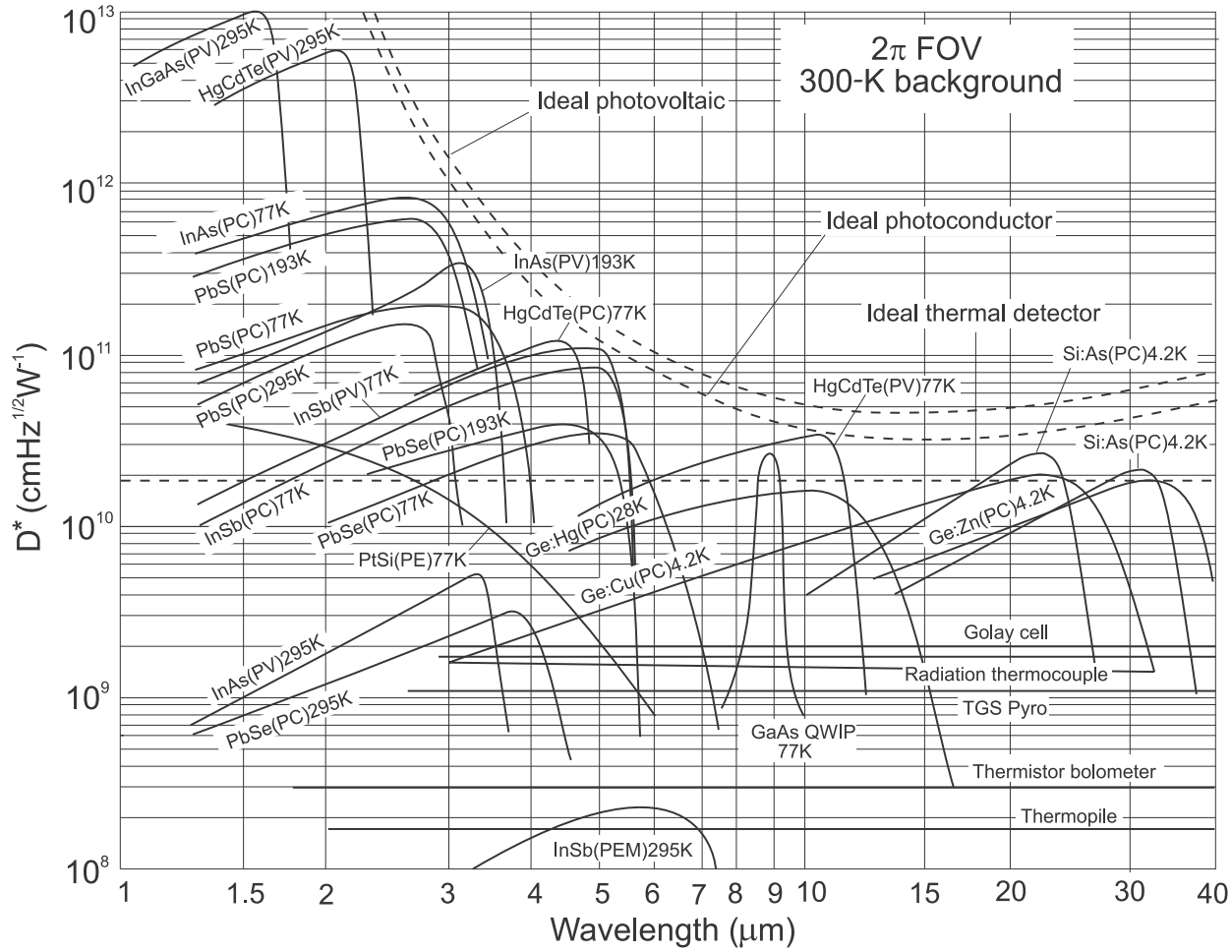


Figure 5.19 Spectral D^* for various detector types. Used with permission from Rogalski.¹²

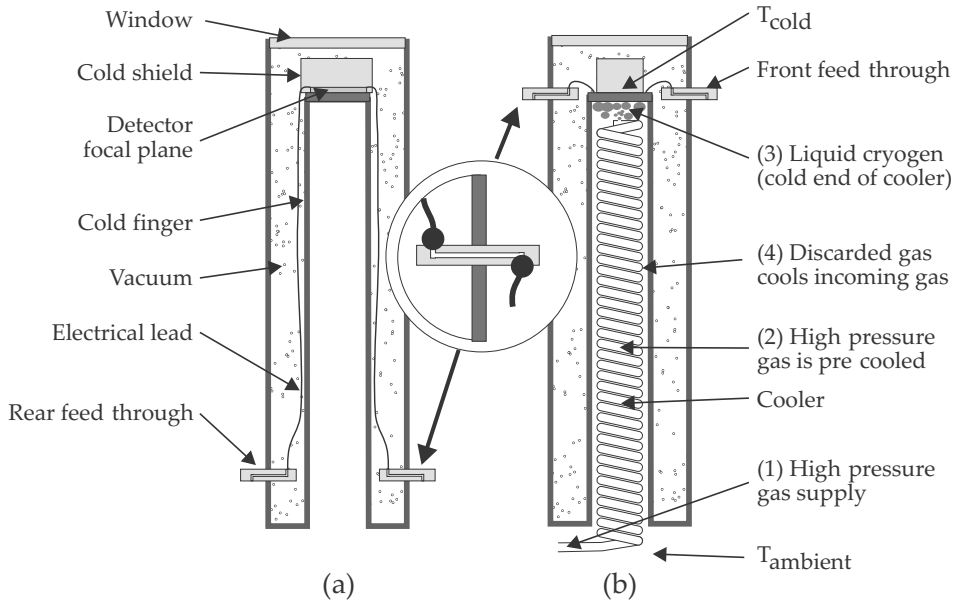


Figure 5.20 Detector vacuum dewar: (a) with wire feed-through and (b) cooler inserted.

Cooling is generally achieved by three techniques: (a) gas/liquid cryogen coolants,⁶⁴ (b) thermo-electric cooling, and (c) radiative cooling. Radiative IR detector cooling is only feasible in space, where the ambient temperature is extremely low. The principle of radiative cooling is energy loss by thermal (Planck-law) radiation from the hot detector to the low background temperature (around 3 K) in space.

Gas/liquid cooling employs a substance called a cryogen, which normally exists as a gas at room temperatures and liquifies at subzero temperatures. The cooling effect of the cryogen is achieved by moving the substance through a series of pressure-volume-temperature (PVT) phase changes, utilizing the appropriate physical process at each of the PVT operating points. There are two types of cryogenic coolers: (a) open-cycle coolers that discard the cryogen after cooling, and (b) closed-cycle coolers that retain the cryogen in a closed circuit of pipes and reservoirs.²⁷ Current research aim to develop micromachined coolers.⁶⁵

Detectors operating at very low temperatures are housed in a mechanical structure called a dewar, shown in Figure 5.20(a). The dewar is, in effect, a small thermos flask with the detector device on the inside cold end of the 'cold finger' (the inside tube of the flask). The dewar is evacuated to very high vacuum in order to minimize the heat load on the cooler. The cooler is inserted into the cold finger and cools down the tip of the cold finger to the detector operating temperature. The detector tempera-

ture is within a few degrees of the cryogen boiling point. Dewars can be made of glass, metal, or a combination of glass and metal. In some cases, the dewar is integrated with the cooler, as a single assembly.

A typical open-cycle Joule–Thomson cooler is shown in Figure 5.20(b). The process (a) starts with a gas at a pressure of 45 MPa at 300 K, and then (b) the gas is pre-cooled by the discarded gas prior to (c) expanding through a very small nozzle, such that the pressure drops to around 1–2 bar, and cools the cryogen to liquification (e.g., 77 K in the case of nitrogen), at which point the liquid cryogen absorbs the heat from the detector and evaporates, (d) flowing away from the detector. The cryogen is selected on the basis of its boiling point and cooling capacity. Popular cryogen gases include helium (4 K), nitrogen (77 K), and argon (87 K), or a mixture of gases.

The Stirling-cycle thermodynamic process is used most commonly in closed-cycle coolers. The cryogen is contained in a sealed piping system and oscillates between the hot and cold ends. The Stirling engine normally employs an electric motor (rotary or linear) to compress the gas. Stirling coolers are available in two configurations: the integral cooler and the split cooler. The integral cooler combines all mechanics and cryogenics into a single unit. The split Stirling cooler separates the compression unit from the cooling unit. The split engine has the advantage of being silent with no vibration coupling from the engine to the detector. This was particularly important during the earlier generations of coolers. Research has resulted in new engines that run much quieter, allowing the integration of the cooling engine with the cooler.

Thermoelectric (TE) cooler modules are solid state heat pumps that employ the Peltier effect — the phenomenon whereby the passage of an electrical current through a junction consisting of two dissimilar metals results in a cooling effect. When the direction of current flow is reversed, heating will occur. A thermoelectric module consists of an array of p- and n-type semiconductor elements heavily doped with electrical carriers, as shown in Figure 5.21. The array of elements is soldered so that it is electrically connected in series and thermally connected in parallel. This array is then fixed to two ceramic substrates; a hot and a cold side. Heat is absorbed at the cold side of the n- and p- type elements. The electrical charge carriers (holes in the p-type; electrons in the n-type) always travel from the cold side to the hot side, and heat is always released at the hot side of thermoelectric element. The temperature differential between the hot and cold ends is inversely proportional to the pump load. At higher heat loads the temperature differential reduces.

The vast majority of thermoelectric coolers achieve temperature differ-

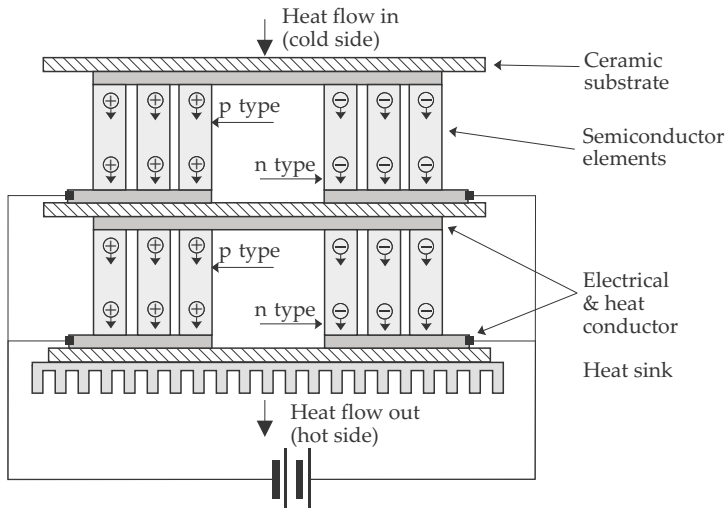


Figure 5.21 Two-stage thermoelectric cooler (adapted⁶⁶).

ences of no more than 50–60 °C across one level of thermoelectric cooling. Multi-stage coolers stack multiple coolers, reaching temperatures down to 195–230 K in a three-stage cooler and down to 50 K with ten stages. As the number of stages increases, the achievable temperature difference increases, but the heat-pumping capacity and overall efficiency decreases. TE coolers are relatively inefficient, resulting in a significant amount of heat that must be dissipated from the hot end. Typical currents range from 1–8 A. Typical supply voltages are 2–15 V. These coolers can cool down heat loads of 0.1–15 W.

5.8 Photoconductive Detectors

5.8.1 Introduction

Photoconductive detectors are photon detectors that change in conductivity upon incident flux. This section derives equations describing the responsivity, frequency response, noise, and D^* of a photoconductive detector. Photoconductive detectors are well documented.^{2,12,13,19,67}

5.8.2 Photoconductive detector signal

The quantum efficiency is given by Equation (5.88), where d is the depth of the detector along the flux propagation direction. From Equation (5.88) it would appear that by increasing the thickness d of the detector, a better quantum efficiency can be obtained. However, when increasing d , the

incremental contribution of the absorbed flux per unit depth decreases. Furthermore, large d also increases the shunt conductance effect of the bulk of the material. The carriers can move freely throughout the volume of the (deeper) detector, resulting in a smaller change in conductance. The optimum detector thickness is approximately $d = 1/\alpha$, achieving a balance between absorption and carrier shunt conductance in the bulk of the material.

The conductivity of the detector bulk material is given by

$$\sigma = q(n_e\mu_e + n_h\mu_h), \quad (5.89)$$

where σ is the conductivity in $[\Omega/\text{cm}]$, q is the electronic charge, n is the carrier density in $[\text{quanta}/\text{cm}^3]$, and μ is the carrier mobility in $[\text{cm}^2/(\text{s}\cdot\text{V})]$. The subscripts e and h denote electron and hole quantities, respectively. Incident photon flux creates free carriers in the material Δn , resulting in a change in conductivity of

$$\Delta\sigma = q(\Delta n_e\mu_e + \Delta n_h\mu_h). \quad (5.90)$$

Photons absorbed in intrinsic semiconductor material result in an equal number of holes and electrons ($\Delta n_e = \Delta n_h$) because the material has relatively high purity with very few traps. In extrinsic detectors, the number of hole carriers and electron carriers are not equal because the extrinsic impurities act as traps for one of the two types of carriers.

The resistance in $[\Omega]$ of the detector is given by

$$R_d = \frac{l}{\sigma w d} = \frac{k}{\sigma}, \quad (5.91)$$

where l is the length of the detector along the direction between the two external electrodes, w is the width, and d is the thickness of the detector; see Figure 5.22. The change in resistance due to photon-excited carriers is

$$\partial R_d = -\frac{k\partial\sigma}{\sigma^2}, \quad (5.92)$$

leading to (for intrinsic detectors)

$$\frac{\partial R_d}{R_d} = -\frac{\partial\sigma}{\sigma} = -\frac{q(\mu_e + \mu_h)\Delta n}{\sigma}, \quad (5.93)$$

where Δn is the number of free photon-excited electron-hole pairs. Δn depends on the change in incoming photon flux $\Delta\Phi_p$ $[\text{q}/\text{s}]$ or irradiance ΔE_q in $[\text{q}/(\text{s}\cdot\text{m}^2)]$, the electron lifetime τ_e , the hole lifetime τ_h , and the detector quantum efficiency η by

$$\Delta n = \frac{\eta\Delta\Phi_p\tau_e}{lwd} = \frac{\eta\Delta E_q\tau_e}{d}, \quad (5.94)$$

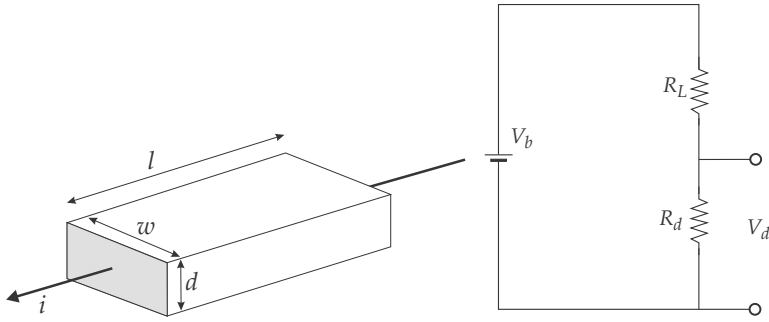


Figure 5.22 Photoconductive detector geometry and bias circuitry.

hence

$$\frac{\partial R_d}{R_d} = -\frac{q(\mu_e \tau_e + \mu_h \tau_h) \eta \Delta \Phi_p}{l w d \sigma}. \quad (5.95)$$

The flux in Equation (5.95) is defined in photon quantities. Assuming monochromatic detector operation, the flux in radiometric terms $\Phi_e = \Phi_p h \nu$ can be substituted to yield the fractional change in resistance

$$\frac{\partial R_d}{R_d} = -\frac{q(\mu_e \tau_e + \mu_h \tau_h) \eta \Delta \Phi_e \lambda}{h c l w d q (n_e \mu_e + n_h \mu_h)} \quad (5.96)$$

$$= -\frac{(\mu_e \tau_e + \mu_h \tau_h) \eta \lambda}{h c l w d (n_e \mu_e + n_h \mu_h)} \Delta \Phi_e, \quad (5.97)$$

where $\Delta \Phi_e$ is the monochromatic incoming flux in [W], n_e and n_h are the carrier densities in [quanta/cm³], and μ_e and μ_h are the carrier mobility in [cm²/(s·V)]. The subscripts e and h denote electron and hole quantities, respectively; $0 \leq \eta \leq 1$ is the detector quantum efficiency; and l , w , and d are the detector dimensions in length, width, and depth.

5.8.3 Bias circuits for photoconductive detectors

The photoconductive detector is biased as indicated in Figure 5.22, with a constant voltage bias supply and a series detector. The bias condition is $i = V_{\text{bias}} / (R_L + R_d)$, $V_d = V_{\text{bias}} R_d / (R_L + R_d)$. Analysis shows that

$$\frac{\partial V_d}{\partial R_d} = -\frac{V_{\text{bias}} R_L}{(R_L + R_d)^2}, \quad (5.98)$$

the voltage responsivity with units [V/W] now becomes

$$\mathcal{R}_v = \frac{\partial V_d}{\partial \Phi_e} = -\frac{V_{\text{bias}} R_L R_d q (\mu_e \tau_e + \mu_h \tau_h) \lambda \eta}{(R_L + R_d)^2 h c l w d \sigma}. \quad (5.99)$$

The optimal choice for R_L is such that $\partial \mathcal{R}_v / \partial R_L = 0$, which requires $R_L = R_d$; then,

$$\mathcal{R}_v = \left(\frac{-R_L}{R_L + R_d} \right) \left(\frac{V_d(\mu_e \tau_e + \mu_h \tau_h)}{\sigma d l w} \right) \left(\frac{q \lambda \eta}{h c} \right) \quad (5.100)$$

$$= \left(\frac{-R_L R_d}{R_L + R_d} \right) \left(\frac{V_d(\mu_e \tau_e + \mu_h \tau_h)}{l^2} \right) \left(\frac{q \lambda \eta}{h c} \right) \quad (5.101)$$

$$= G_c G_{ph} \mathcal{R}. \quad (5.102)$$

The derivation identifies three distinct terms: (a) the bias circuit gain G_c in units of $[\Omega]$ or $[V/A]$, (b) the unitless bulk material photoconductive gain G_{ph} , and (c) the photocurrent responsivity \mathcal{R} in units $[A/W]$. The photoconductive gain can be written

$$G_{ph} = \frac{V_d(\mu_e \tau_e + \mu_h \tau_h)}{l^2} = \frac{\varepsilon(\mu_e \tau_e + \mu_h \tau_h)}{l}, \quad (5.103)$$

where ε is the electric field across the detector in $[V/m]$. $\varepsilon(\mu_e \tau_e + \mu_h \tau_h)$ has units of velocity (carrier velocity), so that $l/[\varepsilon(\mu_e \tau_e + \mu_h \tau_h)]$ represents the transit time the carrier needs to travel along the length of the detector, τ_t . The photoconductive gain is then given by

$$G_{ph} = \frac{\tau}{\tau_t}, \quad (5.104)$$

which is the shortest carrier lifetime divided by the transit time. G_{ph} therefore represents the number of times the carriers can move across the length of the detector within its average lifetime. Typical values vary from 0.5–1.0 for silicon to 10^3 – 10^4 for HgCdTe.

Equation (5.100) shows that the responsivity can be increased by reducing the equilibrium conductivity of the detector. The conductivity can be reduced by reducing the number of free carriers under equilibrium conditions. This can be achieved by cooling the detector and by reducing the incident background flux on the detector.

5.8.4 Frequency response of photoconductive detectors

The frequency response of a photoconductive detector is determined by the carrier lifetime. Ignore the thermally generated carriers and assume that the carrier lifetime is constant and independent of carrier concentration. If the number of carriers under equilibrium conditions (no illumination) is given by n_0 , the variation in the number of carriers due to the optical excitation can be denoted by $\Delta n(t)$ around the nominal n_0 , i.e., $n(t) = \Delta n(t) + n_0$. The change in the variation of optically excited carriers is given by

$$\frac{d\Delta n(t)}{dt} = \eta \Phi_q(t) - \frac{\Delta n(t)}{\tau}, \quad (5.105)$$

where $\eta\Phi_q$ is the optically generated carrier rate, η is the quantum efficiency, τ is carrier lifetime, $\Phi_q(t)$ is the flux incident on the detector in [q/s], and $\Delta n(t)/\tau$ is the rate of recombination of free (optically generated) carriers. Taking the Fourier transform on both sides of Equation (5.105),

$$i2\pi f \Delta N(f) = \eta\Phi_q(f) - \frac{\Delta N(f)}{\tau}, \quad (5.106)$$

$$\Delta N(f) \left(i2\pi f + \frac{1}{\tau} \right) = \eta\Phi_q(f), \text{ and} \quad (5.107)$$

$$\Delta N(f) = \frac{\eta\Phi_q(f)\tau}{1 + i2\pi f\tau}, \quad (5.108)$$

where f is frequency, $\Delta N(f)$ is the Fourier transform of the number of free carriers due to optical excitation, $\Phi_q(f)$ is the Fourier transform of the photon flux incident on the detector, and τ is the carrier lifetime. The frequency response of the detector is then simply

$$H(f) = \frac{\Delta N(f)}{\eta\Phi_q(f)} = \frac{\tau}{1 + i2\pi f\tau}. \quad (5.109)$$

Equation (5.108) describes the frequency response of the photon-to-carrier conversion process, and therefore the frequency response of the detector output signal relative to the incident optical flux. The responsivity [Equation (5.103)] of the detector can be increased by increasing the carrier lifetime. However, increasing the carrier lifetime reduces the high-frequency response of the detector [Equation (5.109)].

5.8.5 Noise in photoconductive detectors

There are four components that contribute to the total noise in the photoconductive detector: Johnson noise, $1/f$ noise, g-r noise due to optical flux, and g-r noise due to thermally excited carriers:¹⁹

$$i_t^2 = 4k \left(\frac{T_d}{R_d} + \frac{T_L}{R_L} \right) + \frac{k_1 I^\alpha}{f^\beta} + 4q^2 \eta \Phi_{qb} G_{ph}^2 + 4g_{\theta} q G_{ph}^2. \quad (5.110)$$

IR photoconductive detectors are normally cooled as explained in Section 5.7. Suppose further that the detector is not used at frequencies where $1/f$ noise dominates, $k_1 I^\alpha / f^\beta \ll i_t^2$, then two performance-limiting noises are present: generation-recombination-noise-limited operation and Johnson-noise-limited operation.

5.8.5.1 Generation–recombination-noise-limited operation

In order to obtain generation–recombination-noise-limited operation, the Johnson noise i_J must be much less than the g-r noise i_{gr} , $i_J \ll i_{gr}$ or

$$4q^2\eta\Phi_{qb}G_{ph}^2 \gg 4k \left(\frac{T_d}{R_d} + \frac{T_L}{R_L} \right). \quad (5.111)$$

The g-r noise does not depend on the detector temperature. Hence, generation–recombination-noise-limited performance can be obtained by reducing the Johnson noise by cooling the detector and load resistor to a temperature $T_L = T_d = T$, where

$$T \ll \frac{q^2G_{ph}^2\eta\Phi_{qb}R_{\text{eff}}}{k}, \quad (5.112)$$

and R_{eff} is the effective parallel resistance of the detector and load resistor. The NEP of the detector in [W] is then (see Section 7.1)

$$NEP_\lambda = \frac{\text{noise}}{\mathcal{R}_v} \quad (5.113)$$

$$= \frac{i_{gr}\Delta f R_{\text{eff}}}{\mathcal{R}_v} \quad (5.114)$$

$$= \frac{2hc}{\lambda} \sqrt{\left(\frac{A_d \Delta f}{\int_0^{\lambda_c} E_q d\lambda} \right)}. \quad (5.115)$$

Then the D^* is given by (see Section 7.1)

$$D^*(\lambda) = \frac{\sqrt{A_d \Delta f}}{NEP} \quad (5.116)$$

$$= \frac{\lambda}{2hc} \sqrt{\frac{\eta}{\int_0^{\lambda_c} E_q d\lambda}}, \quad (5.117)$$

where A_d is the area of the detector. Note that the D^* depends only on the background flux E_q ; this condition is also known as background-limited operation. The dashed line in Figure 5.19 was calculated with Equation (5.117).

5.8.5.2 Johnson-noise-limited operation

The detector is Johnson-noise limited if

$$\frac{T}{R_{\text{eff}}} \gg \frac{q^2G_{ph}^2\eta\Phi_{qb}}{k}, \quad (5.118)$$

where $T_L = T_d = T$, and R_{eff} is the effective parallel resistance of the detector and load resistor. If the load resistor is at the same temperature as the detector $T_L = T_d$,

$$i_J^2 = 4k \left(\frac{T_d}{R_d} + \frac{T_L}{R_L} \right) = \frac{4kT_d}{R_{\text{eff}}}. \quad (5.119)$$

The NEP of the detector in [W] is then

$$\text{NEP}_\lambda = \frac{\text{noise}}{\mathcal{R}_v} \quad (5.120)$$

$$= \frac{i_J \Delta f R_{\text{eff}}}{\mathcal{R}_v} \quad (5.121)$$

$$= \frac{hc}{q\eta\lambda} \left(\frac{\sqrt{4kT_d \Delta f}}{G_{ph} \sqrt{R_{\text{eff}}}} \right), \quad (5.122)$$

and the D^* is given by (see Section 7.1)

$$D^*(\lambda) = \frac{\sqrt{A_d \Delta f}}{\text{NEP}} \quad (5.123)$$

$$= \frac{q\lambda\eta}{2hc} \frac{G_{ph} \sqrt{R_{\text{eff}} A_d}}{\sqrt{kT_d}}. \quad (5.124)$$

The factor $G_{ph} \sqrt{R_{\text{eff}} A_d}$ is not intrinsic to the material it depends on the detector geometry and size. HgCdTe has low R_{eff} but very high gain G_{ph} , resulting in a high D^* .

5.9 Photovoltaic Detectors

5.9.1 Photovoltaic detector operation

A photovoltaic detector is a device where photon-generated carriers are converted to a current and swept out of the device by an internal potential inherent in the device's construction. The device described in this section is a p-n diode, also known as a photodiode.^{2,3,7,10,12,16,19,22,46,67,68} Electron-hole pairs formed as minority carriers in the depletion region or within a diffusion length of the space-charge volume depletion region contribute to the current flow. The photocurrent generated in the p-n diode creates a reverse current flow, shifting the current-voltage curve toward a negative current. Figure 5.23 provides a conceptual illustration of p-n and pin photodiode construction.

If an n-impurity material is doped with p-type dopants (more commonly done), or if very pure (intrinsic) semiconductor material is doped

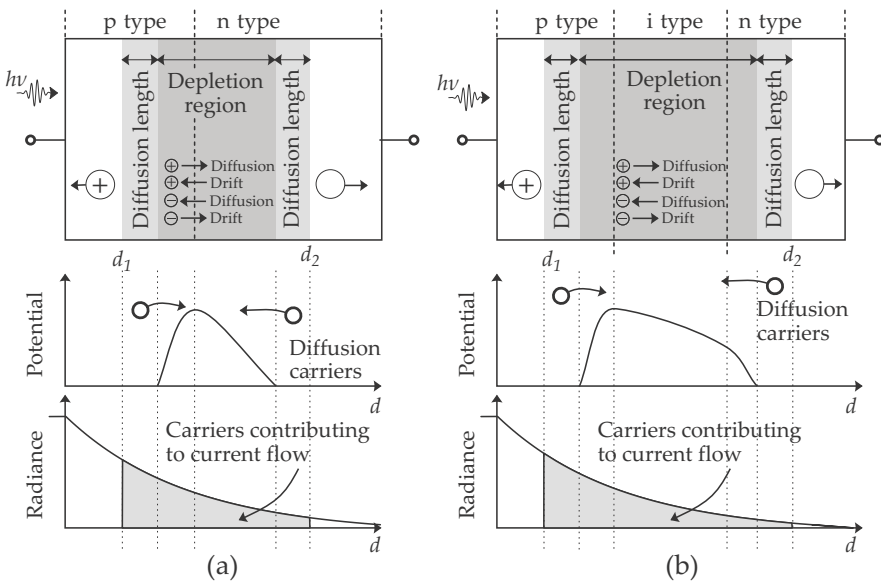


Figure 5.23 Photovoltaic detector construction: (a) p-n diode and (b) pin diode (adapted⁶⁸).

with p-type and n-type dopants (less commonly done), the interface between these two doped regions is called the p-n junction. The n material has an excess of electrons, leaving a positive ion in the crystal lattice; in contrast, the p material accepts an electron from the lattice, leaving a negatively charged ion. The free electrons and holes will diffuse across the junction, leaving a region with positive and negative ions, but with no free charge carriers — this is called the depletion region. There are no free carriers in the depletion region, resulting in an internal field across the depletion region. Real-world detectors are complex devices designed to locate the depletion layer as close to the detector front surface as possible. Various techniques are also used to optimize detector size and responsivity, and to reduce noise. Figure 5.24 shows the electronic and energy state in, and around the p-n junction. Note in particular the concentration profiles along the depth of the p-n junction.⁴⁶

If a photon is absorbed in the depletion region, the resulting electron-hole pair forms minority carriers in the depletion region. These minority carriers are accelerated and swept out of the depletion region under the built-in electric field present across the depletion region. The resultant minority current results in a measurable current on the device's terminals.

The operation of a photovoltaic detector and a photoconductive detector differs in the sense that any carrier pair created in the photoconductive detector contributes to the lower detector resistance. In a photoconductive

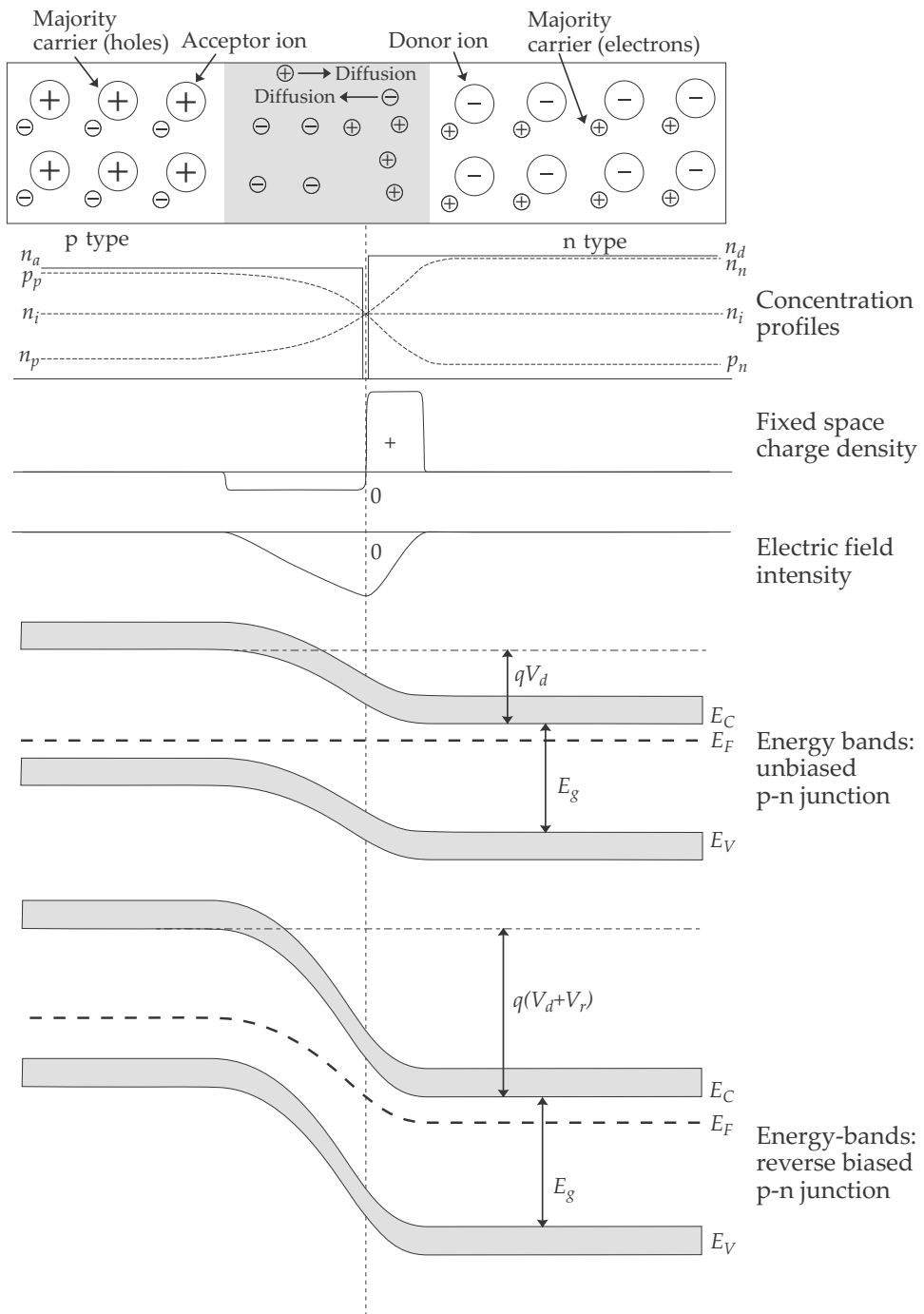


Figure 5.24 The p-n diode junction, energy diagrams, and energy bands.

detector, there is no requirement that the photon be absorbed in any particular spatial region in the detector because an electrical field is applied over the whole detector. In the photovoltaic detector, the field only exists across the depletion region (in the absence of an external bias voltage). The photons must therefore be absorbed *within or near the depletion region*, otherwise the electron-hole pair recombines without ever contributing to the current flow. The diffusion length is a material property that indicates the radius of a sphere beyond which the electron-hole pair will probably recombine within the carrier lifetime. Electrons absorbed within one diffusion length from the depletion region will reach the depletion region and be swept out as signal. For any electron-hole pair created beyond one diffusion length from the depletion region, there is a near-zero probability that the hole or electron will diffuse into the depletion region and will be detected. The diffusion length for holes and electrons are, respectively,

$$L_h = \sqrt{kT\mu_h\tau_h/q} \quad (5.125)$$

$$L_e = \sqrt{kT\mu_e\tau_e/q}, \quad (5.126)$$

where μ is the carrier mobility, and τ is the carrier lifetime.

The spectral responsivity of a photovoltaic detector is described in Equation (5.5). The quantum efficiency shown in Equation (5.88) assumes that all photons absorbed in the material are converted to electrons. This condition is not met in the case of photodiodes because electron-hole pairs formed beyond the diffusion length of the depletion region do not contribute to current flow. For a photovoltaic detector the quantum efficiency is given by

$$\eta = (1 - \rho)(e^{-\alpha d_1} - e^{-\alpha d_2}), \quad (5.127)$$

where d_1 and $d_2 > d_1$ define the depth of the active region (depletion width plus twice the diffusion length). From Equation (5.127) it would appear that by increasing the thickness of the depletion region $d_2 - d_1$ of the detector, a better quantum efficiency can be obtained. This is the prime motivation behind the silicon pin detector, where an intrinsic layer is used to increase the depletion width.

5.9.2 Diode current–voltage relationship

The I-V curve describes the operating point of a diode by relating the voltage across the terminals and the current flowing out of the terminals. All diodes have this behavior (see also Section 9.3.2.4). The diode current is related to the voltage across the device by the nonlinear function:^{2,3,10,12,16}

$$I = I_{\text{sat}}e^{qV/(kT\beta)} - I_{\text{sat}} - I_{ph}, \quad (5.128)$$

where I_{ph} is photocurrent resulting from the absorbed photons [Equation (5.4) or (5.5)], I_{sat} is the reverse-bias-saturation current ($I > 0$ implies forward-bias current), R_0 is the dark resistance, V is the voltage across the diode ($V > 0$ implies forward bias), T is temperature, q is the charge on an electron, k is the Boltzmann constant, and β is the material-dependent ‘nonideality’ factor^{2,12} (with value between 1 and 2). When the diffusion current dominates (e.g., silicon), $\beta = 1$. If the recombination current dominates, $\beta = 2$. The symbol for reverse-bias-saturation I_{sat} is also written as I_0 in some texts.

The first term, $I_{sat}e^{qV/(kT\beta)}$, is the diffusion current, comprising the electrons (holes) in the conduction (valence) band of the n-type (p-type) material that diffuse into the junction with sufficient energy to overcome the potential barrier between the n- and p-type regions. The second term, I_{sat} , results from the electrons (holes) in the p-type (n-type) material that are thermally excited. If these thermally generated electrons (holes) encounter the junction, they are attracted into the n-type (p-type) region, independent of the existence of an applied voltage. Note that the photoinduced current flows in the same direction as the reverse-bias current because the photoinduced carriers are minority carriers (see Figure 5.25).

$I_{sat} = \beta kT / (qR_0)$ is the reverse-saturation current given by^{2,16}

$$I_{sat} = qA_d \left(\frac{n_p D_e}{L_e} + \frac{p_n D_h}{L_h} \right), \quad (5.129)$$

where A_d is the detector junction area in $[m^2]$, n_p is the electron concentration in the p-type material in $[cm^{-3}]$, D_e is the diffusion constant for electrons in $[cm^2/s]$, L_e is the diffusion length for electrons in $[cm]$, p_n is the hole concentration in the n-type material in $[cm^{-3}]$, D_h is the diffusion constant for holes in $[cm^2/s]$, and L_h is the diffusion length for holes in $[cm]$. Using Equations (5.76) and (5.77) leads to

$$I_{sat} = qA_d \left(\frac{n_i^2 D_e}{n_a L_e} + \frac{n_i^2 D_h}{n_d L_h} \right). \quad (5.130)$$

The diffusion length L is related to the diffusion constant D and carrier lifetime τ by $L = \sqrt{D\tau}$, hence

$$I_{sat} = qA_d \left(\frac{n_i^2 L_e}{n_a \tau_e} + \frac{n_i^2 L_h}{n_d \tau_h} \right). \quad (5.131)$$

5.9.3 Bias configurations for photovoltaic detectors

The photovoltaic detector can be biased in three quadrants and five operating conditions. Strong forward bias is applied to diodes to generate

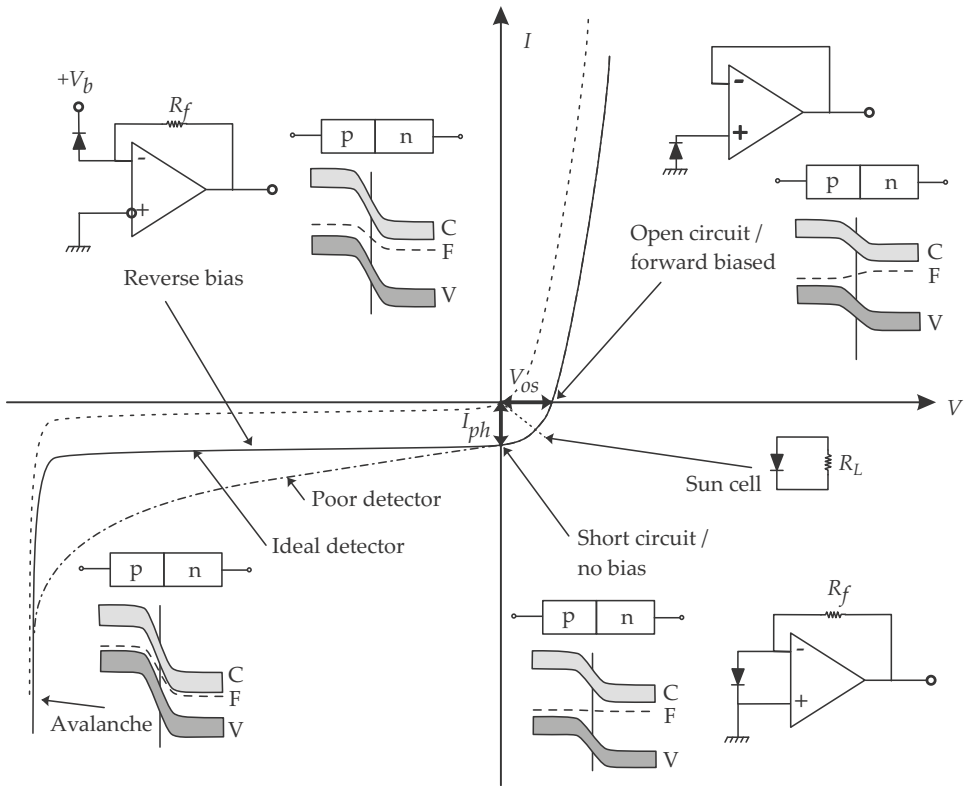


Figure 5.25 Bias configurations and energy bands for various operating conditions.

light (LEDs and laser diodes).^{56,69} This section only considers reverse or small forward bias as indicated in Figure 5.25. All of the bias conditions described in the following sections are commonly found in various applications.

5.9.3.1 Reverse-bias operation of photovoltaic detectors

Under reverse bias, the electron-hole pair generated by the photon forms minority carriers in the depletion region. The hole and electron are swept out by the electric field in the device. All of the electron-hole pairs are externally observed as a current flowing through the device; see Figure 5.26. In Figure 5.26 minority carriers are indicated by small circles, and majority carriers are indicated by large circles. If the reverse-bias voltage is too high, the detector may enter the avalanche regime. Some detectors are designed to operate in the avalanche regime, but special semiconductor and electronics design techniques are required to construct an avalanche device that can operate reliably. The avalanche detector is also very sensitive to bias voltage and temperature variations, and special power-supply

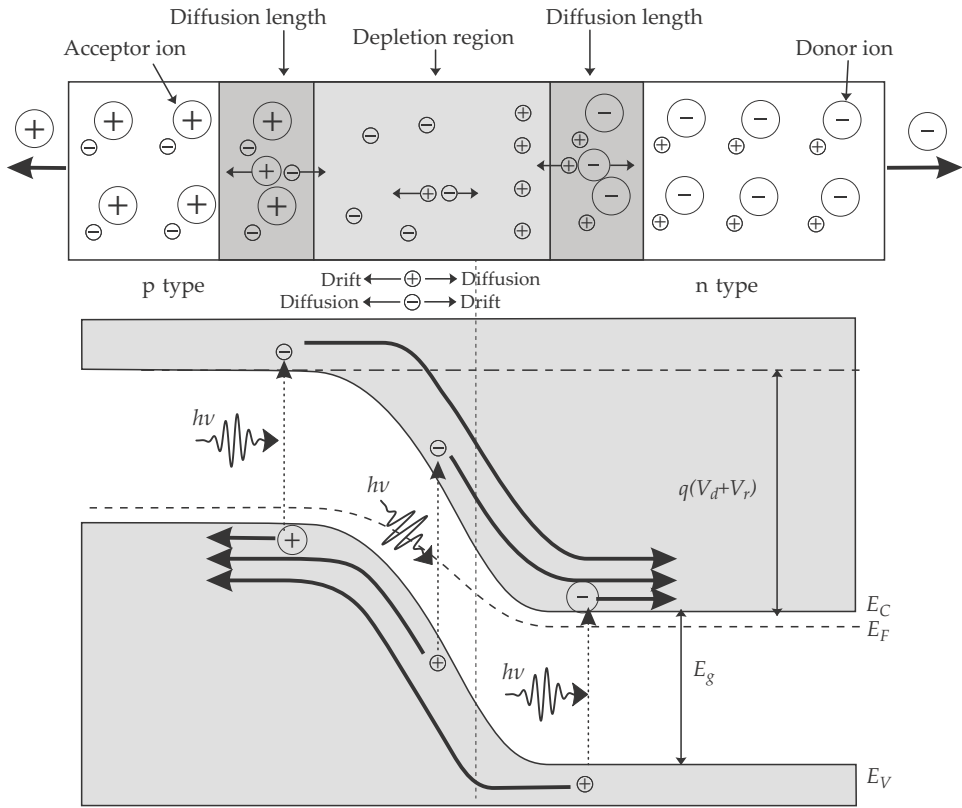


Figure 5.26 The illuminated p-n junction under reverse-bias conditions.

techniques are required to drive an avalanche detector reliably.

The diode current–voltage relationship, under reverse bias $V = -V_{\text{bias}}$, is as follows:

$$I = I_{\text{sat}} e^{-qV_{\text{bias}}/(kT\beta)} - I_{\text{sat}} - I_{ph}. \quad (5.132)$$

If the device reverse-saturation current is well behaved, $e^{-qV_{\text{bias}}/(kT\beta)} \rightarrow 0$ and $I = -I_{\text{sat}} - I_{ph}$. The detector signal is proportional to the photon flux plus the saturation current. The saturation current depends strongly on temperature and greatly influences the detector operation and performance. The reverse-bias mode is commonly used to reduce the depletion region capacitance because an increase in reverse-bias voltage widens the depletion-layer $C = \epsilon A/(d_2 - d_1)$. For a silicon detector the depletion layer capacitance reduces by 70% for every doubling in reverse-bias voltage (for reverse voltages beyond a few volts). The wider depletion layer also results in improved quantum efficiency because a larger volume of the semiconductor is under the depletion-region potential.

The reverse-biased detector electronic circuit model is shown in Fig-

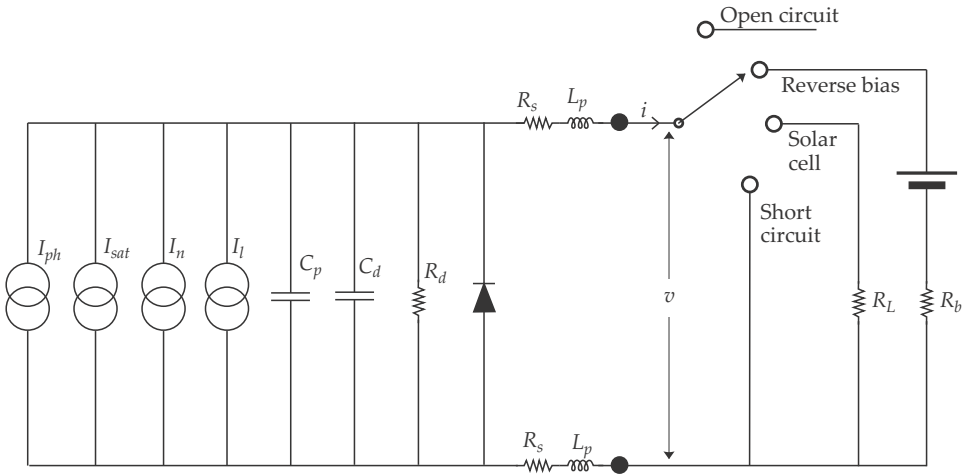


Figure 5.27 Detector circuit model under all of the bias conditions. I_{ph} is the signal current, I_n is the noise current, I_l is the temperature-dependent leakage (dark) current, I_{sat} is the saturation current, C_d is the depletion-layer capacitance, C_p and L_p are the packaging and lead capacitance and inductance, R_d is the detector dynamic resistance (partially forward-biased diode), R_s is the series lead, contact, and bulk resistance, R_b is the detector bias resistance, and R_L is the load resistance.

ure 5.27. Detector data sheets sometimes indicate typical values for these various circuit elements, but the designer must often estimate values for some detector parameters. The detector capacitance depends strongly on the reverse-bias voltage as indicated above. The detector dynamic resistance R_d is equal to the reciprocal of the slope of the I-V curve under reverse bias. Silicon pin detectors have dynamic resistance values of several M Ω . Detectors made from other detector materials, such as InSb, can have much-lower dynamic resistances. The leakage current is the reverse-saturation current. Note that if the detector has a low dynamic resistance, an increase in reverse-bias voltage will also increase the leakage current through the detector. This is not evident in silicon pin detectors but detector performance degrades under high reverse voltage for InSb and HgCdTe detectors.

5.9.3.2 Open-circuit operation of photovoltaic detectors

The photon flux creates minority carriers that are swept out across the depletion layer. However, once out of the depletion layer, the carriers have nowhere to go, and an accumulation of charge takes place. The increase in charge results in an increase of forward voltage across the junction. As the forward voltage across the junction increases, the forward current through the device increases, depleting the accumulation of charge. Under conditions of infinite load impedance, the reverse photocurrent results in

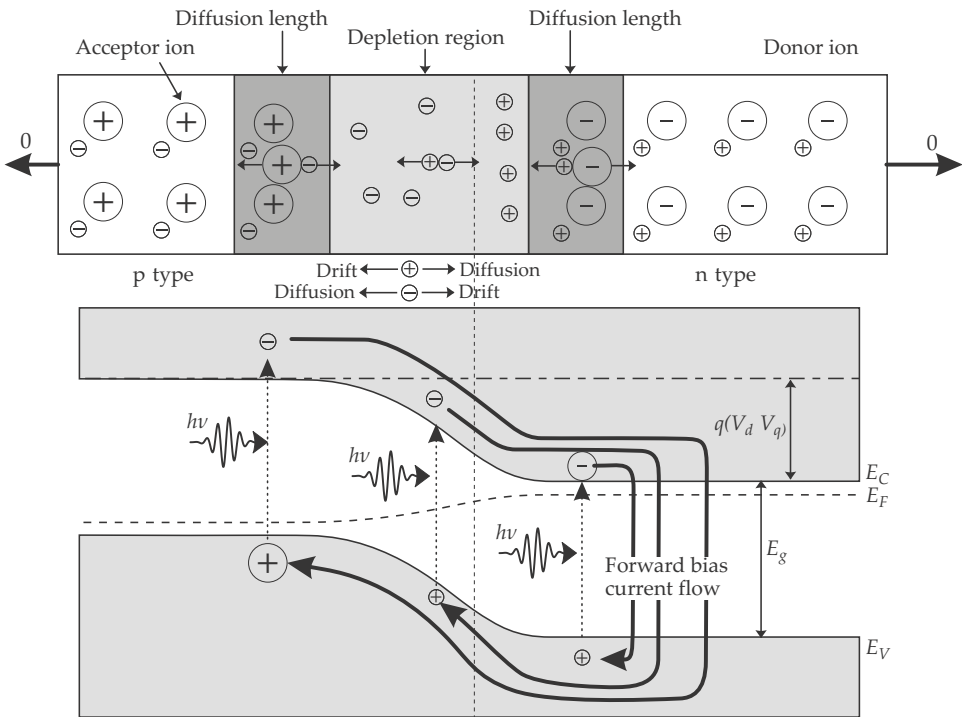


Figure 5.28 The illuminated p-n junction under open-circuit conditions.

exactly the same magnitude forward current, resulting in a zero net output current. Note that in this case the voltage across the diode is equal to the depletion-region barrier voltage. The diode therefore sinks all of the photocurrent under its own voltage; none flows out of the device (see Figure 5.28).

The diode load is a high impedance, such that the detector *external* current approaches zero, $I = 0$, following from Equation (5.128)

$$V = \frac{\beta kT}{q} \ln \left(\frac{I_{ph}}{I_{sat}} + 1 \right) \quad (5.133)$$

$$= \frac{\beta kT}{q} \ln \left(\frac{\eta q \lambda \Phi_e}{hc I_{sat}} + 1 \right). \quad (5.134)$$

The open-circuit output voltage is proportional to the logarithm of the incident flux. This bias mode can therefore be used to build light meters that respond over very wide ranges in a logarithmic fashion.

The detector circuit model for this mode is shown in Figure 5.27. In this mode, the depletion region is very 'thin' resulting in a high depletion capacitance. The detector is therefore *much* slower than under reverse-bias conditions. The detector output voltage is temperature dependent, as can

be seen in Equation (5.134) (most logarithmic converters relying on a diode I-V characteristic are very sensitive to temperature variations). It is very noisy, with mainly flicker noise. The thinner depletion region also results in a smaller quantum efficiency.

5.9.3.3 Optimal power transfer in photovoltaic detectors

Silicon solar cells are operated in a mode to obtain maximum power transfer. The solar cell is allowed to deliver power into a resistor, such that it is neither short-circuit nor open-circuit. The circuit operation is halfway between the open-circuit and short-circuit bias conditions already discussed. A high-value load resistor will cause the internal diode to ‘sink’ the photocurrent internal to the device. On the other hand, a zero-value load impedance means that there is no power delivered into the load.

5.9.3.4 Short-circuit operation of photovoltaic detectors

In the short-circuit bias condition $V = 0$, hence $I = -I_{ph}$. The detector current is directly proportional to the photocurrent and therefore directly proportional to the photon flux absorbed in the depletion region. If the detector has high linearity due to high dynamic resistance (for example, a silicon-pin diode), this bias condition can be used to construct an accurate optical flux meter (called a radiometer or optical flux meter).

The detector circuit model for this mode is shown in Figure 5.27. Note that in this case the voltage across the diode is zero, and the diode has a high dynamic resistance. The dynamic resistance can be in the $M\Omega$ ranges for a silicon-pin or small IR detector, but considerably lower for large IR detectors. The capacitance of the depletion layer is high because the depletion layer is not as wide as it can be under reverse bias.

5.9.4 Frequency response of a photovoltaic detector

The frequency response of a photovoltaic detector is determined by two factors: (1) the time taken by the carriers to move through the depletion layer and (2) (primarily) by the detector capacitance and load-resistance time constant.

If the number of carriers *moving through the depletion layer* under equilibrium conditions, with no illumination, is given by n_0 , the variation in the number of carriers due to the optical excitation can be denoted by $\Delta n(t)$ around the nominal n_0 , i.e., $n(t) = \Delta n(t) + n_0$. The change or variation

due to optically excited carriers is given by

$$\frac{d\Delta n(t)}{dt} = \eta\Phi_q(t) - \frac{\Delta n(t)}{\tau}, \quad (5.135)$$

where $\eta\Phi_q$ is the optically generated carrier rate, η is the quantum efficiency, $\Phi_q(t)$ is the flux incident on the detector in [q/s], and $\Delta n(t)/\tau$ is the rate of carriers moving through the depletion layer. Taking the Fourier transform on both sides of Equation (5.135),

$$i2\pi f\Delta N(f) = \eta\Phi_q(f) - \frac{\Delta N(f)}{\tau} \quad (5.136)$$

$$\Delta N(f) \left(i2\pi f + \frac{1}{\tau} \right) = \eta\Phi_q(f) \quad (5.137)$$

$$\Delta N(f) = \frac{\eta\Phi_q(f)\tau}{1 + i2\pi f\tau}, \quad (5.138)$$

where f is frequency, $\Delta N(f)$ is the Fourier transform of the number of free carriers due to optical excitation, $\Phi_q(f)$ is the Fourier transform of the photon flux incident on the detector, and τ is the time taken for the carrier to move through the depletion layer. The frequency response of the detector is then

$$H(f) = \frac{\Delta N(f)}{\eta\Phi_q(f)} = \frac{\tau}{1 + i2\pi f\tau}. \quad (5.139)$$

How long does the carrier take to move through the depletion layer? Consider a silicon diode with the following characteristics. The carrier mobilities for carriers in silicon are 500 (holes) and 1300 (electrons) $\text{cm}^2/(\text{V}\cdot\text{s})$. A typical silicon-pin detector depletion width is 50 μm under an applied 20-V reverse voltage. The carrier transit time can be calculated from the above information to be of the order of 1 ns. This silicon detector should be able to internally respond to signals well into hundreds of MHz and, with some care, even into the GHz frequencies.

The operational speed of most photovoltaic detectors are limited by their operation in real-life capacitive electronic circuits. The depletion region capacitance, together with the series resistance and load resistance, provides a very tough design challenge if high operating speeds are required.

5.9.5 Noise in photovoltaic detectors

The photovoltaic detector creates shot noise by virtue of its potential barrier across the depletion layer, and Johnson noise in its dynamic resistance.

The dynamic resistance at zero-bias voltage is given by

$$R_0 = - \left(\frac{dI}{dV} \right)^{-1} \bigg|_{V=0} = \frac{\beta kT}{qI_{\text{sat}}}. \quad (5.140)$$

The Johnson noise is then

$$i_n^2 = \frac{4kT}{R_0} = \frac{4qI_{\text{sat}}}{\beta} = \frac{2(2qI_{\text{sat}})}{\beta}, \quad (5.141)$$

which has the same form as shot noise! Recall that the current in the photovoltaic detector is given by

$$I = I_{\text{sat}} e^{qV/(kT\beta)} - I_{\text{sat}} - I_{ph}. \quad (5.142)$$

The first term is the diffusion current, the second term is the thermally generated current, and the third term is due to the photocurrent. Each of these terms is an independent current and creates noise statistically independent from the other currents and must be added vectorially.² The shot noise PSD in [A^2/Hz], created in the detector, is therefore given by

$$i_n^2 = 2qI_{\text{sat}} e^{qV/(kT\beta)} + 2qI_{\text{sat}} + 2qI_{ph}. \quad (5.143)$$

Using Equation (5.140) it follows that

$$i_n^2 = 2q \left[\frac{\beta kT}{qR_0} e^{qV/(kT\beta)} + \frac{\beta kT}{qR_0} + \Phi_q \right], \quad (5.144)$$

where β is the diode nonideal factor, T is temperature in [K], R_0 is the dynamic resistance under zero-bias conditions, and Φ_q in [q/s] is the background flux falling in on the detector, given by

$$\Phi_q = \int_0^{\lambda_c} \eta E_q(T_b) A_d d\lambda, \quad (5.145)$$

where the integral is calculated over the detector's sensitive spectral range, η is the detector quantum efficiency, A_d is the detector area in [m^2], and E_q is the background photon irradiance in [$\text{q}/(\text{m}^2 \cdot \text{s})$] for a thermally radiating background at temperature T_b in [K]. The NEP of the detector in [W] is then

$$NEP_\lambda = \frac{\text{noise}}{\mathcal{R}_\lambda} \quad (5.146)$$

$$= \frac{hc \sqrt{i_n^2 \Delta f}}{\eta q \lambda} \quad (5.147)$$

$$= \frac{hc \sqrt{\Delta f}}{\eta q \lambda} \sqrt{2q \left[\frac{\beta kT}{qR_0} e^{\frac{qV}{kT\beta}} + \frac{\beta kT}{qR_0} + \eta q \Phi_q \right] + \frac{4kT}{R_0}}. \quad (5.148)$$

The detector D^* is given by

$$D_\lambda^* = \frac{\sqrt{A_d \Delta f}}{NEP_\lambda} \quad (5.149)$$

$$= \frac{\eta q \lambda \sqrt{A_d \Delta f}}{hc \sqrt{i_n^2 \Delta f}} \quad (5.150)$$

$$= \frac{\eta q \lambda \sqrt{A_d \Delta f}}{hc \sqrt{2q \Delta f} \sqrt{\frac{\beta kT}{q^2 R_0} e^{qV/(kT\beta)} + \frac{\beta kT}{q R_0} + \eta q \Phi_q + \frac{2kT}{q R_0}}} \quad (5.151)$$

$$= \frac{\eta \lambda}{hc \sqrt{2} \sqrt{\frac{\beta kT}{q^2 A_d R_0} e^{qV/(kT\beta)} + \frac{\beta kT}{q^2 A_d R_0} + \eta E_q + \frac{2kT}{q^2 R_0 A_d}}} \quad (5.152)$$

5.9.5.1 Background-limited operation of photovoltaic detectors

If the noise induced by the background flux exceeds the combined thermal-generation current noise, diffusion current noise, and Johnson noise,

$$\eta E_q \gg \frac{\beta kT}{q^2 A_d R_0} e^{qV/(kT\beta)} + \frac{\beta kT}{q^2 A_d R_0} + \frac{2kT}{q^2 R_0 A_d}, \quad (5.153)$$

the D^* is given by

$$D^* = \frac{\eta \lambda}{hc \sqrt{2} \sqrt{\eta \int_0^{\lambda_c} E_q d\lambda}} = \frac{\lambda}{hc \sqrt{2} \sqrt{\int_0^{\lambda_c} E_q d\lambda}}. \quad (5.154)$$

Compare this equation with Equation (5.117) for the photoconductive detector. The background-limited D^* for a photoconductive and photovoltaic detector differs only by a factor of $1/\sqrt{2}$. Equation (5.154) can be used to determine the upper-limit D^* that can be expected from any detector for a given background. Figure 5.29 shows the theoretical limit for an ideal detector with given cutoff wavelength λ_c against a background of given temperature. In practice, all real detectors fall short of this theoretical prediction.

5.9.5.2 Detector-limited operation: short-circuit mode

Consider a diode with zero-bias voltage $e^{qV/(\beta kT)} \rightarrow 1$. If the noise induced by the background is less than the internal detector noise

$$\frac{2kT}{q^2 R_0 A_d} + \frac{2\beta kT}{q^2 R_0 A_d} \gg \eta \int_0^{\lambda_c} E_q d\lambda, \quad (5.155)$$

then the D^* is given by

$$D^* = \frac{\eta \lambda}{hc \sqrt{2} \sqrt{\frac{2\beta kT}{q^2 R_0 A_d} + \frac{2kT}{q^2 R_0 A_d}}} = \frac{q \eta \lambda \sqrt{A_d R_0}}{2hc \sqrt{(\beta + 1)kT}}. \quad (5.156)$$

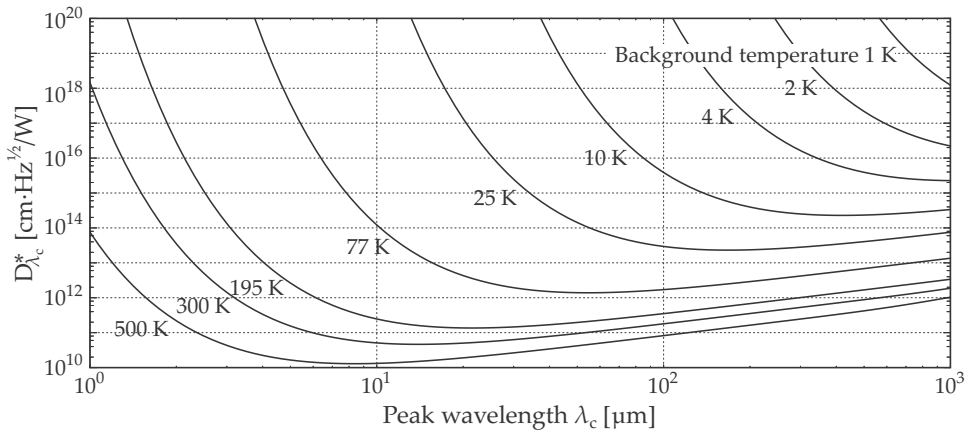


Figure 5.29 Theoretical limits to D^* for given detector cutoff wavelength and background temperatures.

Equation (5.156) indicates that for a given temperature T and quantum efficiency η , the detector D^* is determined by $R_0 A_d$, where A_d is the detector area, and R_0 is the zero-bias dynamic resistance. The $R_0 A_d$ product is an indication of the dark current flowing through the detector and therefore of the noise due to the dark current. Compare Equation (5.156) with Equation (5.124) for a photoconductive detector.

At large reverse-bias voltage the potential barrier is increased, reducing the number of diffusion carriers crossing the junction. It would appear that the noise can be reduced by applying a large reverse-bias voltage;

$$\frac{\beta k T}{q^2 A_d R_0} e^{-qV/(kT\beta)} + \frac{\beta k T}{q^2 A_d R_0} \rightarrow \frac{\beta k T}{q^2 A_d R_0}. \quad (5.157)$$

In practice, this never materializes. At high reverse-bias voltage, the noise increases due to $1/f$ noise in I_{sat} . Most demanding, low-noise, IR-detector applications are designed to operate at near-zero-bias voltage. One exception to this rule is high-performance silicon pin diodes, where the increased reverse-bias voltage has only a small effect on increasing the noise, but significantly reduces the depletion layer capacitance.

5.9.5.3 Detector-limited operation: open-circuit mode

Under open-circuit bias conditions the detector drives an infinitely high impedance, and there is no external current flow. This implies that the internal diode must sink all of the photocurrent: the signal current flows twice through the junction, firstly during generation, and secondly during sinking. These two currents are statistically independent so that the noise power adds. Hence, the open-circuit bias configuration increases the

noise by $\sqrt{2}$ over the short-circuit configuration. The D^* is derived from Equation (5.154) as

$$D^* = \frac{\lambda}{2hc} \sqrt{\frac{\eta}{\int_0^{\lambda_c} E_q d\lambda}}. \quad (5.158)$$

Equation (5.158) confirms that the open-circuit bias mode is noisier than the short-circuit bias mode.

5.9.6 Detector performance modeling

This section demonstrates the development of a simple detector model suitable for high-level system modeling as considered in this book. The performance of the detector is described in terms of its figures of merit. Most of these figures of merit are described in Sections 7.1.2 and 5.3.11. A model is developed for a single-element InSb photovoltaic detector. The equations used and material parameters are detailed in the following description. The material parameters are listed in Appendix A. The current-voltage and spectral results are shown in Figure 5.30. The data and code for this model are available on the pyradi website.⁷⁰

The steps followed are as follows:

1. Calculate the bandgap E_g of the InSb material at the detector operating temperature (80 K), using Equation (5.3), with values $A = 6 \times 10^{-4}$ and $B = 500.0$ (Table A.6). The resulting bandgap is 0.233 eV.
2. The spectral absorption coefficient is calculated by using Equations (5.78) and (5.80) and the data in Table A.6. The spectral absorption coefficient is shown in Figure 5.30(a).
3. The quantum efficiency is calculated by using Equations (5.87) and (5.88), with a refractive index of 3.42 for the detector and 1.0 for air, a depletion layer thickness of 5 μm , and the spectral absorption coefficient calculated in the previous step. The spectral quantum efficiency is shown in Figure 5.30(b).
4. The spectral responsivity is calculated using the spectral quantum efficiency and Equation (5.5). The gain of the photovoltaic detector is 1.0. The spectral responsivity is shown in Figure 5.30(c).
5. Calculate the spectral photon and radiant irradiance on the detector for a target blackbody source with unity emissivity and temperature $T_s =$

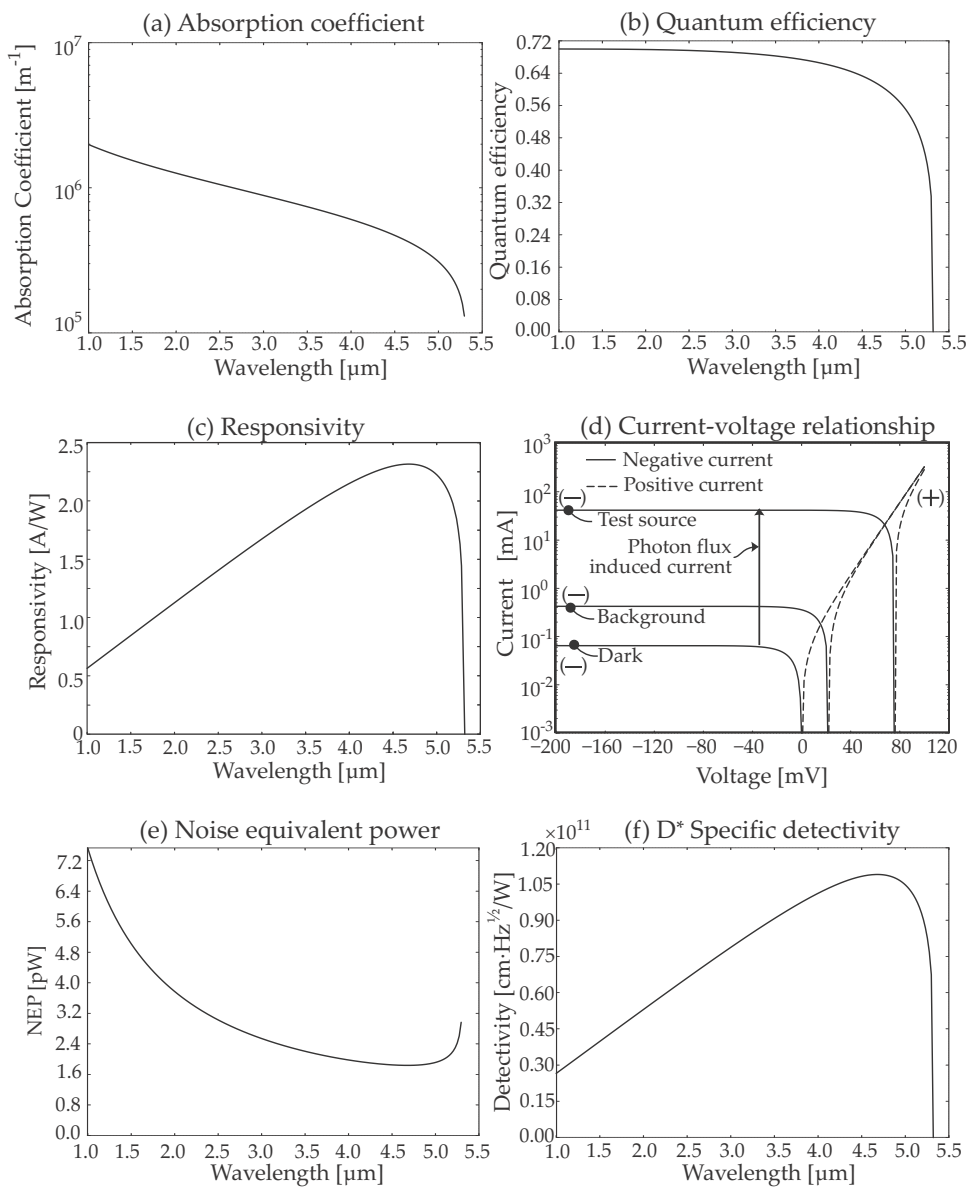


Figure 5.30 InSb detector model results: (a) spectral absorption coefficient, (b) quantum efficiency, (c) spectral responsivity, (d) current–voltage relationship, (e) spectral NEP, and (f) spectral specific detectivity.

2000 K. The source area $A_s = 33 \text{ mm}^2$, located a distance $R = 100 \text{ mm}$ from the detector. The irradiance is calculated as

$$E_{s\lambda} = \epsilon_\lambda M_\lambda(T_s) \omega / \pi = M(T_s) A_s / (\pi R^2), \quad (5.159)$$

where the spectral exitance is given by Equations (3.1) and (3.3) for radiant and photon exitance, respectively.

6. Calculate the spectral photon and radiant irradiance on the detector for a hemispherical background blackbody source with unity emissivity and temperature of 280 K. The calculation is done as for item 5 above but for an hemispherical solid angle.
7. Calculate the current through the detector for source and background irradiance using a simplified version of Equation (6.17):

$$i_{\mathcal{R}} = A_d \hat{\mathcal{R}} \int_0^\infty E_{s\lambda} \mathcal{R}_\lambda d\lambda, \quad (5.160)$$

using the spectral responsivity calculated in item 4 above.

8. Calculate I_{sat} . First, calculate the carrier diffusion lengths L_e and L_h by using Equations (5.125) and (5.126). Second, calculate the intrinsic carrier concentration n_i using Equation (5.74). Equations (5.76) and (5.77) the donor concentration, is used to determine the concentration of carriers. Finally, calculate the reverse-saturation current with Equation (5.131).

The values used in this calculation for the InSb detector are as follows: electron mobility^{7,12} $100.0 \text{ cm}^2/(\text{V}\cdot\text{s})$; electron lifetime¹² $1 \times 10^{-8} \text{ s}$; hole mobility¹² $1.0 \text{ cm}^2/(\text{V}\cdot\text{s})$; hole lifetime¹² $1 \times 10^{-8} \text{ s}$; electron effective mass 0.014 of electron mass; hole effective mass 0.43 of electron mass; acceptor concentration $1 \times 10^{16} \text{ m}^{-3}$; donor concentration $1 \times 10^{16} \text{ m}^{-3}$; E_g as calculated in item 1 above; detector temperature of 80 K; and detector area $200 \text{ }\mu\text{m}$ by $200 \text{ }\mu\text{m}$.

9. Calculate the current–voltage relationship using Equation (5.128). The saturation current I_{sat} is calculated as explained in item 8 above. The photocurrent is determined from Equation (5.4) or (5.5). The value of β is determined as explained in Section 5.9.2. The value for $\beta = 1.7$ is used here because InSb operation is dominated by g-r current.¹²

The current–voltage relationship can be calculated for different background flux values, such as ‘dark’ ($I_{ph} = 0$), or for different irradiance values, such as those calculated in items 5 and 6 above.

Figure 5.30(d) shows the current–voltage relationship for three cases: (a) the dark current (i.e., no background flux), (b) hemispherical background flux at 280 K, and (c) the target flux. Note that the y axis is

logarithmic to show detail — negative values are shown positive because the logarithm of a negative number is not defined.

10. Equation (5.140) is used to calculate the detector's dynamic resistance under zero-bias voltage, in this case using the reverse-saturation current calculated in item 8 above. Equation (5.141) calculates the Johnson noise. Equation (5.143) calculates the noise PSD for the three current components: (a) diffusion, (b) thermal excitation, and (c) photocurrent. Adding all four components in quadrature provides the noise in the detector.

In the example calculation, the value for R_0 was found to be 180 k Ω . Over the 100-Hz noise bandwidth, the Johnson noise is 1.563×10^{-12} A, the shot noise is 3.956×10^{-12} A, and the total noise is 4.254×10^{-12} A.

11. The NEP is determined from Equation (5.25), with the spectral responsivity calculated in item 4 and the noise current calculated in item 10. The NEP is shown in Figure 5.30(e).
12. The D^* is calculated from the area of the detector, the noise bandwidth, the noise, and the responsivity, by Equation (5.26). The responsivity is shown in Figure 5.30(f).

5.10 Impact of Detector Technology on Infrared Systems

The requirements for future detector technology are driven by convenience needs (cost, size, weight, and power) as well as new capabilities (multicolor capability, larger arrays).⁷¹ In order to meet these future needs, current technologies are optimized incrementally and new technologies are being developed.⁸ High-performance IR imaging is the single-most dominant driver in current research. In the medium term, system designers can expect a wide range of detector products, from top-end large multicolor arrays to low-end inexpensive thermal detectors. Several technology status reviews are available.^{8,27,41,72}

The cost of high-performance IR detector arrays precludes their widespread use in consumer electronics. Recent years have seen meteoric advances in uncooled detector array technology, both in performance improvement and cost reduction. Microelectromechanical systems (MEMS) technology provided the means to construct thermal detector elements with low heat capacity and short time constants. Some of these detector arrays achieve 0.02–0.05-K NETD below 100-Hz noise bandwidth. Thermal detector elements achieve D^* values of 10^8 to 10^9 cm $\cdot\sqrt{\text{Hz}}/\text{W}$.⁸ Focus is

now on lowering cost in quantity production. The number of detector elements per chip follows Moore's law (doubling the number of pixels every 18 months). The availability of these low-cost IR detector arrays opens up new applications in the consumer market.

Detector cooling requirements are major cost, size, weight, and power drivers. Traditional technologies are incrementally improved to support operation at higher temperature. A few decades ago, long-wave photon detectors had to be cooled to 4 K. Present-day technology requires long-wave detectors to operate at 77–85 K. In the near-term future the requirement for low operating temperatures will be relaxed. Background-limited operation can be achieved at a detector temperature of 100–120 K for LWIR detectors and 160 K for MWIR detectors.^{7,8}

Array sizes are ever increasing with megapixel IR staring array detectors readily available at competitive prices. Array sizes in conventional technology are approaching the limit of practical element size — larger arrays will require larger detectors and concomitant scale increase in optics. Large optics represent exponential increase in system cost. Hence, several new technologies are investigated to reduce pixel size. Staring array detectors in the LWIR are subject to limited charge accumulation in the read-out electronics — a particular problem in the LWIR because of the large photon flux at the longer wavelengths. This remains a challenge to be resolved by new capacitor technologies.

HgCdTe is a very versatile material, supporting spectral band optimization and multicolor arrays. The material is subject to several issues related to a weak Hg-Te bond, but many of these have been resolved. None of the newer materials yet provide fundamental advantages over the established HgCdTe technology base, which is expected to remain as a material of choice in many detector solutions. HgCdTe technology is mature in the MWIR but less so in the LWIR spectral band. InAsSb can be tuned to provide LWIR spectral coverage, with similar predicted performance.⁸ In the shorter-wavelength bands there are many new material types, offering high performance, low noise, and tunable alloys. The InGaAs alloy composition can be adjusted to enable detectors from 1.6–2.6 μm . $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ provides a viable alternative to Ge.

QWIP detectors rely on bandgap alteration, such that new energy states are created with bandgaps optimized for the required spectral range. QWIP detectors are manufactured in GaAs/AlGaAs for MWIR and LWIR applications. HgCdTe still outperforms QWIP detectors in terms of spectral range and D^* . QWIP detectors are more uniform than HgCdTe detectors, which eases the burden of nonuniformity correction. QWIP detectors also lend themselves better to very large arrays. QWIP detectors are less

susceptible to high background photon flux (but also less sensitive to target signals) because the detector spectral width is narrower than HgCdTe.

Multicolor detectors are currently in advanced research, and high-performance detectors should appear on the market in the short-to-medium term. One approach is to build a HgCdTe shorter-wavelength detector on top of a longer-wavelength detector. Another approach uses monolithic structures with different spectral responses as already reported in QWIP IR photodetectors.^{73,74}

Several new technologies are being investigated, including quantum dot detectors,⁷⁵ nearly matched lattice detectors,⁷⁶ and strained-layer superlattice detectors.⁷⁷ Strained-layer superlattice detectors have bandgaps smaller than any of the constituent materials. Quantum dot detectors confine carriers in three-dimensional space using nanostructures. These technologies are still in the research phase.

Bibliography

- [1] Singh, J., *Electronic and Optoelectronic Properties of Semiconductors Structures*, Cambridge University Press, Cambridge, UK (2003).
- [2] Dereniak, E. L. and Boreman, G. D., *Infrared Detectors and Systems*, John Wiley & Sons, New York (1996).
- [3] Rogalski, A., *Infrared Detectors*, 1st Ed., Gordon and Breach Science Publishers, Amsterdam, The Netherlands (2000).
- [4] Rogalski, A., "Infrared Detectors: an overview," *Infrared Physics & Technology* 43, 187–210 (2002).
- [5] Hamamatsu, "Characteristics and use of infrared detectors," Technical report SD-12, Hamamatsu (2004).
- [6] Rogalski, A., "Infrared Photovoltaic Detectors," *Opto-electronics Review* 5, 205–216 (1997).
- [7] Piotrowski, J. and Rogalski, A., *High-Operating-Temperature Infrared Photodetectors*, SPIE Press, Bellingham, WA (2007) [doi: 10.1117/3.717228].
- [8] Rogalski, A., "Infrared detectors: status and trends," *Progress in Quantum Electronics* 27, 59–210 (2003).
- [9] Vurgaftmann, I., Meyer, J. R., and Ram-Mohan, L. R., "Band Parameters for III-V Compound Semiconductors and their Alloys," *Journal of Applied Physics* 89(11), 5815–5875 (2001).

- [10] Piprek, J., *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*, Academic Press, San Diego, CA (2003).
- [11] Murdin, B., Adams, A., and Sweeney, S., "Band Structure and High-pressure Measurements," Ch. 2 in *Mid-infrared Semiconductor Optoelectronics*, Krier, A., Ed., 93–126, Springer, Berlin (2006).
- [12] Rogalski, A., *Infrared Detectors*, 2nd Ed., CRC Press, Boca Raton, FL (2011).
- [13] Palmer, J. M. and Grant, B. G., *The Art of Radiometry*, SPIE Press, Bellingham, WA (2009) [doi: 10.1117/3.798237].
- [14] van der Ziel, A., *Noise in Solid State Devices and Circuits*, Wiley, New York (1986).
- [15] van der Ziel, A., *Noise: Sources, Characterization, Measurement*, Prentice-Hall (1970).
- [16] Gowar, J., *Optical Communication Systems*, Prentice-Hall, Upper Saddle River, NJ (1984).
- [17] Kingston, R. H., *Optical Sources, Detectors, and Systems: Fundamentals and Applications*, Academic Press, San Diego, CA (1995).
- [18] Kingston, R. H., *Detection of Optical and Infrared Radiation*, Springer, Berlin (1978).
- [19] Keyes, R. J., *Optical and Infrared Detectors*, Springer-Verlag, Berlin (1980).
- [20] Mandelbrot, B. B., *Multifractals and $1/f$ noise: Wild Self-Affinity in Physics (1963–1976)*, Springer, Berlin (1999).
- [21] D'Agostino, J. A. and Webb, C. M., "Three-dimensional analysis framework and measurement methodology for imaging system noise," *Proc. SPIE* **1488**, 110–121 (1991) [doi: 10.1117/12.45794].
- [22] Ready, J., "Optical Detectors and Human Vision," Ch. 6 in *Fundamentals of Photonics*, Roychoudhuri, C., Ed., SPIE Press, Bellingham, WA (2008) [doi: 10.1117/3.784938].
- [23] Nyquist, H., "Thermal Agitation of Electric Charge in Conductors," *Physical Review* **32**(1), 110–113 (1928) [doi: 10.1103/PhysRev.32.110].
- [24] Budzier, H. and Gerlach, G., *Thermal Infrared Sensors: Theory, Optimisation and Practice*, John Wiley & Sons, New York (2011).

- [25] Kruse, P. W., *Uncooled Thermal Imaging Arrays, Systems, and Applications*, SPIE Press, Bellingham, WA (2001) [doi: 10.1117/3.415351].
- [26] Datskos, P. and Lavrik, N., "Uncooled Infrared MEMS Detectors," Ch. 12 in *Smart Sensors and MEMS*, Yurish, S. Y. and Gomes, M. T. S. R., Eds., 381–430, Kluwer Academic Publishers, Dordrecht, The Netherlands (2003).
- [27] Rogalski, A., "Progress in focal plane array technologies," *Progress in Quantum Electronics* 36, 342–473 (2012).
- [28] Biberman, L., Ed., *Electro-Optical Imaging: System Performance and Modeling*, SPIE Press, Bellingham, WA (2001).
- [29] Vollmerhausen, R. H., Reago, D. A., and Driggers, R. G., *Analysis and Evaluation of Sampled Imaging Systems*, SPIE Press, Bellingham, WA (2010) [doi: 10.1117/3.853462].
- [30] Bielecki, Z., "Readout electronics for optical detectors," *Optoelectronics Review* 12(1), 129–137 (2004).
- [31] Scott, L. B. and Agostino, J. A., "NVEOD FLIR92 Thermal Imaging Systems Performance Model," *Proc. SPIE* **1689**, 194–203 (1992) [doi: 10.1117/12.137950].
- [32] King, K. F., Bernstein, M. A., and Zhou, X. J., *Handbook of MRI Pulse Sequences*, Elsevier Academic Press, London, UK (2004).
- [33] Rogatto, W. D., Ed., *The Infrared and Electro-Optical Systems Handbook: Electro-Optical Components*, Vol. 3, ERIM and SPIE Press, Bellingham, WA (1993).
- [34] Li, C., Skidmore, G., Howard, C., Han, C., Wood, L., Peysha, D., Williams, E., Trujillo, C., Emmett, J., Robas, G., Jardine, D., Wan, C.-F., and Clarke, E., "Recent development of ultra small pixel uncooled focal plane array at DRS," *Proc. SPIE* **6542**, 65421Y (2007) [doi: 10.1117/12.720267].
- [35] Niklaus, F., Vieider, C., and Jakobsen, H., "MEMS-Based Uncooled Infrared Bolometer Arrays — A Review," *Proc. SPIE* **6836**, 68360D, Bellingham, WA (2007) [doi: 10.1117/12.755128].
- [36] Song, W.-B. and Talghader, J. J., "Design and characterization of adaptive microbolometers," *J. Micromech. Microeng* 16(5), 1073–1079 (2006) [doi: 10.1088/0960-1317/16/5/028].

- [37] Tissot, J., Trouilleau, C., Fieque, B., Crastes, A., and Legras, O., "Un-cooled microbolometer detector: recent developments at Ulis," *Proc. SPIE* **5957**, 59570M (2005) [doi: 10.1117/12.621884].
- [38] Marshall, D. E., "A Review of Pyroelectric Detector Technology," *Proc. SPIE* **132**, 110–117 (1978) [doi: 10.1117/12.956063].
- [39] Vandermeiren, W., Stiens, J., Shkerdin, G., Kotov, V., Tandt, C. D., and Vounckx, R., "Infrared Thermo-Electric Photodetectors," Ch. 8 in *Laser Pulse Phenomena and Applications*, Duarte, F. J., Ed., 143–164, Intech Open, New York (2010).
- [40] Kasap, S. O., *Principles of Electronics Materials and Devices*, 3rd Ed., McGraw-Hill, New York (2006).
- [41] Rogalski, A., "HgCdTe infrared detector material: history, status and outlook," *Reports on Progress in Physics* **68**, 2267–2336 (2005).
- [42] Ashcroft, N. W. and Mermin, N. D., *Solid State Physics*, Harcourt College Publishing, Fort Worth, TX (1976).
- [43] Sze, S. M. and Ng, K. K., *Physics of Semiconductor Devices*, 3rd Ed., Wiley-Interscience, New York (2007).
- [44] Yu, P. Y. and Cardona, M., *Fundamentals of Semiconductors: Physics and Materials Properties*, Springer, Berlin (1996).
- [45] Rezende, S. M., *Materiais e Dispositivos Eletrônicos*, Livraria da Física, São Paulo (2004).
- [46] Ibach, H. and Lüth, H., *Solid-State Physics: An Introduction to Principles of Material Science*, 4th Ed., Springer, Berlin (2009).
- [47] Jackson, K. A. and Schröter, W., *Handbook of Semiconductor Technology*, Wiley-VCH, Berlin (2000).
- [48] Rockett, A., *The Materials Science of Semiconductors*, Springer, Berlin (2008).
- [49] Tilley, R., *Understanding Solids*, John Wiley & Sons, New York (2004).
- [50] Neamen, D. A., *Semiconductor Physics and Devices*, 3rd Ed., McGraw-Hill, Boston (2003).
- [51] Nag, B. R., *Physics of Quantum Wells Devices*, Kluwer Academics Press, Dordrecht, The Netherlands (2001).

- [52] Yacobi, B. G., *Semiconductor Materials: An Introduction to Basic Principles*, Kluwer Academic Publishers, Dordrecht, The Netherlands (2004).
- [53] Fox, M., *Optical Properties of Solids*, Oxford University Press, Oxford, UK (2001).
- [54] Schneider, H. and Liu, H. C., *Quantum Well Infrared Photodetectors*, Springer, Berlin (2007).
- [55] Dresselhaus, M. S., "Solid State Physics (Four Parts)," <http://web.mit.edu/afs/athena/course/6/6.732/www/texts.html>.
- [56] Schubert, E. F., *Light-Emitting Diodes*, Cambridge University Press, Cambridge, UK (2003).
- [57] Hecht, E., *Optics*, 4th Ed., Addison Wesley, Boston, MA (2002).
- [58] Chur, J. and Sher, A., *Physics and Properties of Narrow Gap Semiconductors*, Springer, Berlin (2008).
- [59] Huag, H. and Koch, S. W., *Quantum theory of the Optical and Electronic Properties of Semiconductors*, 4th Ed., World Scientific, London, UK (2004).
- [60] Toyozawa, Y., *Optical Processes in Solids*, Cambridge University Press, Cambridge, UK (2003).
- [61] Bauer, G., "Determination of Electron Temperatures and of Hot Electron Distribution Functions in Semiconductors," Ch. 1 in *Solid-State Physics*, Höhler, G., Ed., 1–106, Springer, Berlin (1974).
- [62] Vasileska, D., "Physical and Mathematical Description of the Operation of Photodetectors," <http://nanohub.org/resources/9142>.
- [63] Rogalski, A., Antoszewski, J., and Faraone, L., "Third-generation infrared photodetector arrays," *Journal of Applied Physics* 105(9), 1–44 (2009) [doi: 10.1063/1.3099572].
- [64] Horn, S. B., "Cryogenic cooling options for forward looking infrared (FLIR)," *Proc SPIE* 0245 (1980) [doi: 10.1117/12.959339].
- [65] Lerou, P.-P. P. M., *Micromachined Joule–Thomson cryocooler*, PhD thesis, University of Twente, Enschede (February 2007).
- [66] TE Technology, "Thermoelectric Coolers: FAQ & Technical Information," <http://www.tetech.com/FAQ-Technical-Information.html>.

- [67] Boyd, R. W., *Radiometry and the Detection of Optical Radiation*, John Wiley & Sons, New York (1983).
- [68] Quimby, R. S., *Photonics and Lasers: An Introduction*, Wiley-Interscience, New York (2006).
- [69] Ossicini, S., Pavesi, L., and Priolo, F., *Light Emitting Silicon for Microphotonics*, Springer, Berlin (2003).
- [70] Pyradi team, "Pyradi data," <https://code.google.com/p/pyradi/source/browse>.
- [71] Daniels, A., *Infrared Systems, Detectors and FPAs*, 2 Ed., SPIE Press (2010).
- [72] Rogalski, A., "Material considerations for third generation infrared photon detectors," *Infrared Physics and Technology* 50, 240–252 (2007).
- [73] Alves, F. D. P., Santos, R. A. T., Amorim, J., Issmael, A. K., and Karunasiri, G., "Widely Separate Spectral Sensitivity Quantum Well Infrared Photodetector Using Interband and Intersubband Transitions," *IEEE Sensors Journal* 8(6), 842–848 (2008).
- [74] Santos, R. A. T., Alves, F. D. P., Taranti, C. G. R., Filho, J. A., and Karunasiri, G., "Quantum wells infrared photodetectors: design using the transfer matrix method," *Proc. SPIE* 7298, 72980B (2009) [doi: 10.1117/12.820251].
- [75] Martyniuk, P. and Rogalski, A., "Quantum-dot Infrared Photodetectors: status and outlook," *Progress in Quantum Electronics* 32(3–4), 89–120 (2008).
- [76] Ting, D. Z., Hill, C. J., Soibel, A., Nguyen, J., Keo, S. A., Lee, M. C., Mumolo, J. M., Liu, J. K., and Gunapala, S. D., "Antimonide-based barrier infrared detectors," *Proc. SPIE* 7660, 76601R (2010) [doi: 10.1117/12.851383].
- [77] Razeghi, M., Hoffman, D., Nguyen, B.-M., Delaunay, P.-Y., Huang, E. K., Tidrow, M. Z., and Nathan, V., "Recent Advances in LWIR Type-II InAs/GaSb Superlattice Photodetectors and Focal Plane Arrays at the Center for Quantum Devices," *Proceedings of the IEEE* 97, 1056–1066 (2009) [doi: 10.1109/JPROC.2009.2017108].
- [78] Pyradi team, "Pyradi Radiometry Python Toolkit," <http://code.google.com/p/pyradi>.

Problems

- 5.1 What choice of bias resistance R_L yields optimal detectivity for a photoconductive detector? Three different approaches can be used: maximum power in the load, maximum change in power in the load, or maximum responsivity. What do you think should be the optimal choice for R_L , given a detector resistance of R_d ? [3]
- 5.2 Ignoring the material properties, calculate the spectral responsivity values for a photovoltaic detector at wavelengths of 0.5, 1, 2.5, 5, and 10 μm for quantum efficiencies of 0.5 and 1.0. [2]
- 5.3 Repeat the previous calculation for the detector materials listed in the following table. In the table the material energy gap is given in electron volts. One electron volt (1 eV) is the amount of energy that an electron acquires when it is accelerated through a potential difference of 1 volt. 1 eV is therefore 1.6×10^{-19} joule.

Material	E_g	Material	E_g
Cds	2.4	PbS	0.42
CdSe	1.8	PbSe	0.23
GaAs	1.35	InSb	0.23
Si	1.12	HgCdTe	0.1
Ge	0.67		
	eV		eV

- 5.4 This is an extension of the previous problems: draw the spectral response of the various detectors accurately to scale on the same graph, ranging from 0–14 μm . There is no need to do extensive calculations and plots, just plot a few key points on the graph. Use Matlab[®] or Python[™] to get numerical solutions. [4]
- 5.5 Do a conceptual design of a two-layer sandwich detector that must cover the spectral region from 2.5–14 μm . Describe which materials must be used in each layer. Which layer must be in front? Draw the spectral response of the combined sandwich, for each of the two layers, on the same graph. Note that the two detectors are electrically independent but not optically independent! [6]
- 5.6 Calculate the response time of a silicon-pin diode with capacitance of 5 pF, series resistance of 100 Ω , and negligible inductance and leakage current in each of the four bias modes. [4]
- 5.7 Calculate the transit time of carriers through the silicon detector described in Section 5.9.4. [2]

- 5.8 Verify Equation (5.143). Make sure that you understand why the noise currents seem to add together. [2]
- 5.9 Determine the spectral responsivity of a silicon detector over the spectral range 0.25–1.2 μm . Use the spectral absorption coefficient data on the pyradi website.⁷⁸ To simplify the calculation you may approximate the absorption coefficient by two or three straight-line segments between 0.25–1.2 μm .
- Consider the following two depletion layer designs: (1) $d_1 = 30 \mu\text{m}$, $d_2 = 60 \mu\text{m}$, and (2) $d_1 = 4 \mu\text{m}$, $d_2 = 34 \mu\text{m}$. In both cases the depletion depth is 30 μm , but in the first case the depletion region is deep inside the detector, whereas the second is close to the surface of the detector.
- Which design is more suited to the detection of IR laser pulses at 0.9 μm ? Which design is more suited to the detection of blue or ultraviolet laser pulses at 0.33 μm ? How would you design a silicon detector for optimum responsivity at a particular wavelength? [8]
- 5.10 Compare the equations for D^* between the photoconductive and photovoltaic cases. [2]
- 5.11 Use Equations (5.145) and (5.154) (or derive your own) to confirm any three points in Figure 5.29. [6]
- 5.12 Use the spectral absorption coefficient data on the pyradi website⁷⁸ to determine the coefficients for all of the materials, fitting Equations (5.78), (5.79), and (5.80). [10]
- 5.13 Review Section 9.7 and then recalculate Figure 5.29 for a conical background FOV with an half-apex angle of 15 deg. The cold shield is at 77 K. [10]
- 5.14 A popular tea-time discussion topic involves the timing of when to add the milk to the tea. Given a time lapse T , which cup will be warmer: the cup with milk added at $t = 0$, or the cup with milk added at $t = T$? Develop a mathematical model for the problem and find an answer. Ignore heat conduction through the base of the cup (the tea is served in styrofoam cups). Assume a constant heat capacity of 4.15 kJ/(K·k). Plot the temperature of the tea versus time for three cases: (a) milk at $t = 0$, (b) milk at $t = T$, and (c) no milk. [8]
- 5.15 Derive Equation (5.34) in full, with motivation for the formulation and derivation. [2]
- 5.16 Design a bolometer thermal detector element for use in an image array operating at a frame rate frequency of 20 Hz. The detector

element will be used in an array of 512×512 elements with size $100 \mu\text{m} \times 100 \mu\text{m}$. Explain how your design optimizes performance. [4]

- 5.17 Photon emission from a blackbody follows the Poisson probability distribution. Calculate the SNR (see Section 7.1.3) inherent in blackbody radiation as a function of temperature.
- 5.18 Show by mathematical derivation and by plotting the result how a cold shield will increase a detector's D^* by $1/\sin \theta$. [6]
- 5.19 Calculate and plot the thermal-fluctuation-noise PSD over the frequency range 10^0 to 10^4 Hz, for $C = 4 \times 10^{-9}$ J/K, $G = 10^{-7}$ W/K at temperatures of 100 K, 300 K, and 500 K. What is the time constant? Comment on the results. [4]
- 5.20 Use Equation (5.56) to calculate and plot the D^* of a band-limited thermal detector for wavelengths from 1 to $10^3 \mu\text{m}$. Consider (a) no filter, (b) only the lower-wavelength cutoff, (c) only the longer-wavelength cutoff, and (d) cutoff at both wavelengths. Also plot the D^* for a photon detector with no filter. Comment on your results. [6]
- 5.21 Develop a mathematical model for the internal quantum efficiency of a photoconductive detector as a function of depth of the device. Determine the thickness of the detector in terms of the absorption coefficient for an internal quantum efficiency of 50% and 95%. Implement the model in software and plot the internal quantum efficiency as a function of detector depth. [6]
- 5.22 Derive an equation for photon detector responsivity in terms of photon flux. Do a detailed dimensional analysis and investigate the units of the photon-flux responsivity. [2]
- 5.23 Calculate and plot the spectral responsivity for a silicon, a germanium, an InSb, and a HgCdTe detector. [6]

Calculate the spectral D^* for the above detectors, for background limited operation when (1) immersed in 300 K background and (2) facing vertically up, looking through a 300-K atmosphere into space. The detector is equipped with a cold shield (Section 9.7) with half-apex angle of 10 deg. Comment on your observations. [6]

The data for this problem is given in the DP04.zip data file on the pyradi website.⁷⁰

Chapter 6

Sensors

The soul never thinks without a picture.
Aristotle

6.1 Overview

This chapter provides an introductory overview of sensors. The analysis is limited to small-angle (paraxial) optics. The purpose is to equip the reader to do a first-order design in the system context of this book. Detailed sensor design is beyond the scope of this chapter and indeed not required for this text. Fundamental to the sensor concept is the geometry of solid angles and how these are effected in the sensor. The second important element is the conversion of optical energy into electrical energy, including the effect of noise.

The path by which a ray propagates through an optical system can be mathematically calculated. The sine, cosine, and tangent functions are used in this calculation. These functions can be written as infinite Taylor series, i.e., $\sin(x) = x - x^3/3! + x^5/5! - x^7/7! + \dots$. The paraxial approximation only uses the first term in the sum $\sin(x) \approx \tan(x) \approx x$, $\cos(x) \approx 1$. The paraxial approximation is valid only for rays at small angles with and near the optical axis. The paraxial approximation can be effectively used for first-order design and system layout despite the small-angle limitations.

The coverage of detectors and noise in this chapter is, similarly, only a brief introduction: sufficient detail is given to support first-order design and modeling.

6.2 Anatomy of a Sensor

The electro-optical system performs two functions. First, the optics focus the 'object' (the thing looked at) into the 'image' (a geometrically scaled representation of the object).¹⁻⁶ Second, the sensor collects and converts

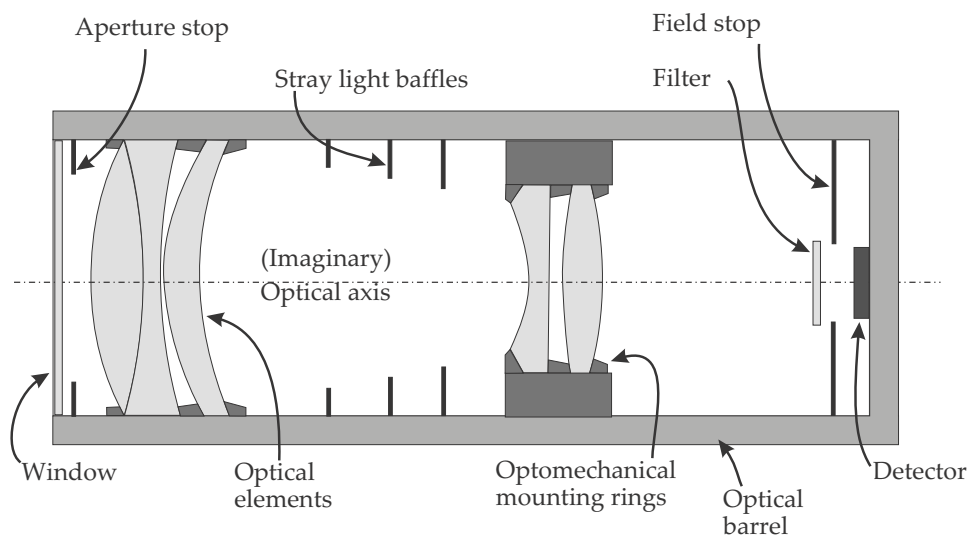


Figure 6.1 Anatomy of a sensor.

optical energy into a useable signal for subsequent use. In the system chain, the sensor is the third building block, after the source and medium. Figure 6.1 shows a conceptual representation of an optical sensor, defining some terminology for use in this chapter.

The optical flux enters from the left through a window. The window is normally a transparent optical material; e.g., BK7 glass for visual-band systems or ZnSe for infrared systems. The window will probably have a thin, multilayer anti-reflection coating to reduce optical reflections from the window surfaces (see Section 3.4.3). Infrared window materials may have a hard coating to protect the soft material of the window against scratches.

The aperture stop is often designed as a mechanical device located such that vignetting is reduced or eliminated. Some systems do not have a dedicated aperture stop, and the diameter of one of the lenses performs the function.

The optical elements (windows, lenses, and mirrors) can be divided into two categories: elements with optical power and elements without optical power. Surfaces without power (plano surfaces) have no focusing effect, whereas surfaces with power (nonplanar surfaces) affect the divergence or convergence of optical rays. The optical elements are held in place in the optical barrel by mounting rings, locating the elements accurately in separation and lateral displacement. Some form of strain relief is required to prevent the optical elements from breaking at temperature extremes — a typical strain relief is a rubber spacer on a noncritical surface.

Optical elements can also be bonded together or glued to metal mounting structures.

Stops or baffles are sometimes used to prevent or minimize internal stray light reflections. The baffles are normally made of thin metal plates spaced at regular intervals. These baffles are covered in an absorbing coating to absorb rather than reflect the stray light. A commonly-used device called a light trap consists of two absorbing plates at a small angle with each other. Any light entering the trap reflects down the two plates, being absorbed at each reflection (see Section 3.2.5).

Some optical sensors have spectral filters that selectively spectrally transmit the optical flux. The filter, together with the detector spectral response, defines the spectral response of the sensor. At its opaque wavelengths, a filter emits thermal radiation, as does any other object.

The field stop, if present, can be placed at any location in the optical system where an image is formed (see the next section). For simple lenses, this is typically near the back focal plane. For systems with internal focal planes (e.g., the Gregorian telescope at the bottom of Figure 6.10), the field stop can be located at such an internal focal plane. If a detector is used, the boundary of the active detector area effectively defines the field stop because the detector has zero responsivity outside the detector active area.

6.3 Introduction to Optics

6.3.1 Optical elements

Optical elements include lenses, mirrors, windows, prisms, filters, and similar items. Of particular interest here are optical elements with optical power. The optical power in a curved optical element refracts or reflects the optical rays so as to converge or diverge the rays. Optical power can be used to form an image of an object in a plane called the focal plane. Figure 6.2 shows the key characteristics of a lens. Assuming unity index of refraction in the object and image planes, the lens equation relates the object properties and image properties with the following equations:

$$\frac{1}{s'} = \frac{1}{s} + \frac{1}{f'} \quad (6.1)$$

$$x'x = -f'^2, \text{ and} \quad (6.2)$$

$$m = \frac{h'}{h} = \frac{s'}{s} = \frac{\sin \theta}{\sin \theta'} \quad (6.3)$$

where s' is the image distance, s is the object distance, f is the focal length, h' is the image height, and h is the object height. The distances s and s' are positive measured toward the right and negative measured toward the left (the same as x in the Cartesian coordinate system).

The optical axis is the 'centerline' of the optical system, normally defining zero field angle. Optical systems have 'planes' along the optical axis with special properties. Figure 6.2 shows a system with two principal planes P_1 and P_2 , and two focal planes F_1 and F_2 . These planes are ideally plano (flat) and perpendicular to the optical axis. In real-world optics, these 'planes' are not flat but could be any shape, depending on the design. In a simple optical system the principal and focal planes tend to be near-spherical. More-sophisticated systems are designed to achieve near-flat focal planes because detector surfaces are flat. The planes are often rotationally symmetric and centered on the optical axis, but this depends on the design of the system. In this chapter, the term 'plane' is used to denote an optical plane (which is normally not mathematically flat).

A ray passing through the first F_1 (front) or second F_2 (back) focal points will be parallel to the optical axis after traversing the lens. A ray passing through the first or second principal points P_1 or P_2 will traverse through the lens with no change in direction. An object at location H_1 will be imaged to the image location at H_2 (in the paraxial approximation).

The location of H_1 (at distance s) and H_2 (at distance s') are called conjugates of each other. 'Infinite conjugates' is a special but common case: when the object is located at infinity $s \rightarrow -\infty$, then $s' = f$, and the image is formed in the back focal plane of the lens or telescope. Most real-world optical systems operate at near-infinite conjugates because the object is normally located a substantial distance from the optical system.

The marginal ray is the ray from the optical axis at the object or image, through the edge of the pupil. The chief ray is the ray from the edge of the field stop, through the principal points of the optics. The chief ray propagates through the lens with no change in direction. The field angle is defined by the chief ray, as the angle defined the image height h' and s' or the object height h and s .

A real optical system has the first and second principal points at separate locations, as shown in Figure 6.2(a). For first-order studies the optical system is often simplified to the 'thin-lens' approximation shown in Figure 6.2(b). In the thin-lens approximation the front and rear principal points coincide. The principal points and focal points all lie in planes, perpendicularly intersecting the optical axes at P , F_1 , and F_2 .

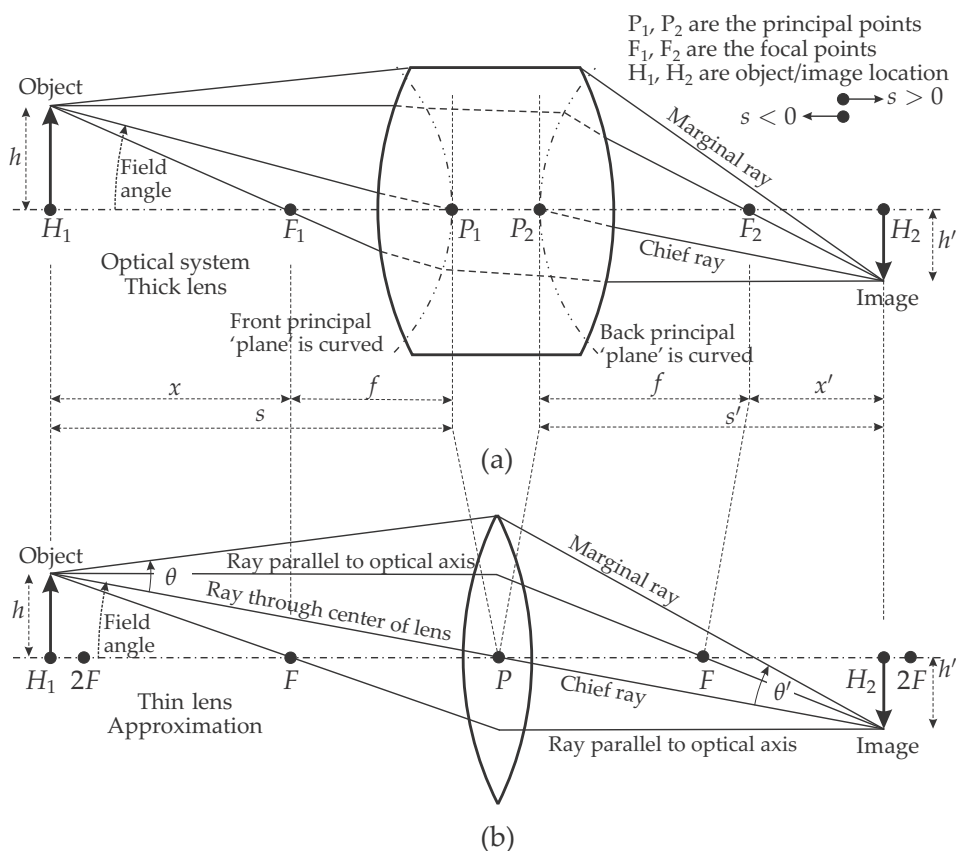


Figure 6.2 Object and image relationship in an image: (a) thick lens and (b) thin-lens approximation.

6.3.2 First-order ray tracing

Figure 6.2(a) shows the imaging relationship in a complex lens. In the paraxial (thin-lens) approximation, the optical system is simplified as shown in Figure 6.2(b). Consider now a few applications of first-order paraxial ray tracing through the optical system. Figure 6.3 shows only three cases, there are many more. From Equation (6.1) it follows that if $s = -2f$, then $s' = 2f$ and $m = -1$, leading to the top picture in Figure 6.3. This means that an object placed at twice the focal length of a lens will image to twice the focal length behind the lens, with a magnification of 1, but form an inverted image.

Increasing to $s = -3f$ leads to $s' = 3f/2$ and $m = -0.5$. Note the trend: for (negatively) increasing s , the image moves closer to F , and the magnification decreases. In the limit for $s \rightarrow -\infty$, $s' \rightarrow f$ and $m \rightarrow 0$. Hence, as the object moves further away from the lens, the image ap-

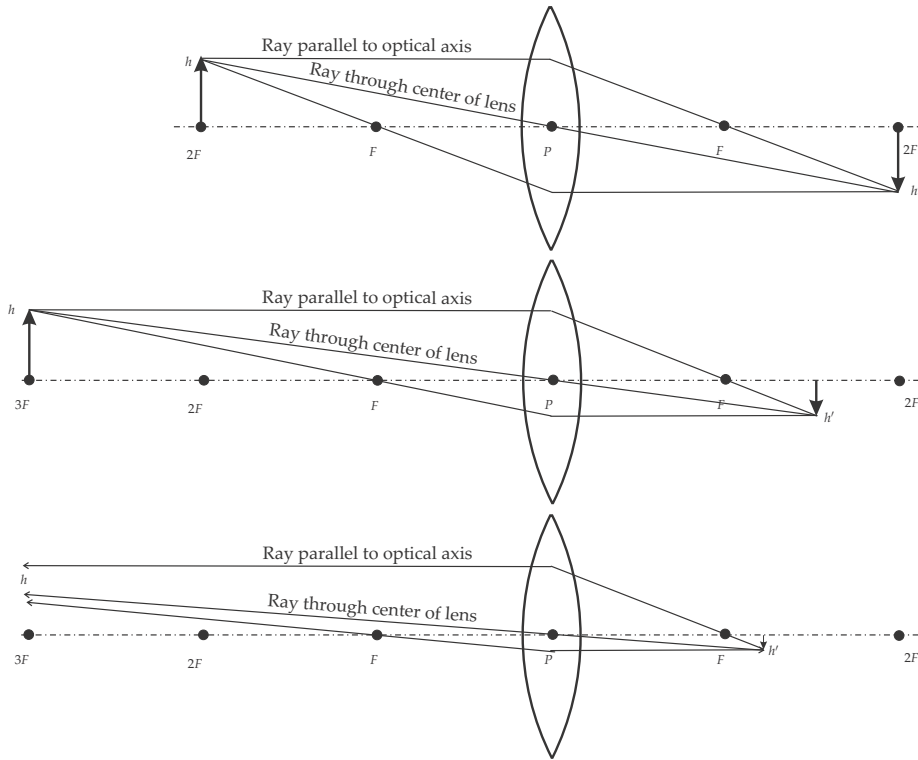


Figure 6.3 First-order ray tracing.

proaches the back focal plane F of the lens.

The results in Figure 6.3 can be constructed geometrically by two simple rules: first, a ray entering a lens parallel to the optical axis always passes through the rear focal plane of the lens. Second, a ray passing through the center (principal points) of the lens does not change its direction. Any point on the object can be mapped by these rules to a point on the image. Likewise, the inverse also applies.

The application of these simple rules leads to the sensor model definition shown in Figure 6.4, where the detector is located in the focal plane of the lens, imaging an object at infinity. The sensor field of view (FOV) is defined by the two chief rays through the center of the lens.

6.3.3 Pupils, apertures, stops, and f -number

An optical aperture, pupil, or a stop is a well-defined opening in an optical system with the purpose of limiting the amount of light, to reject stray light, or to clearly define the field of view. This well-defined opening

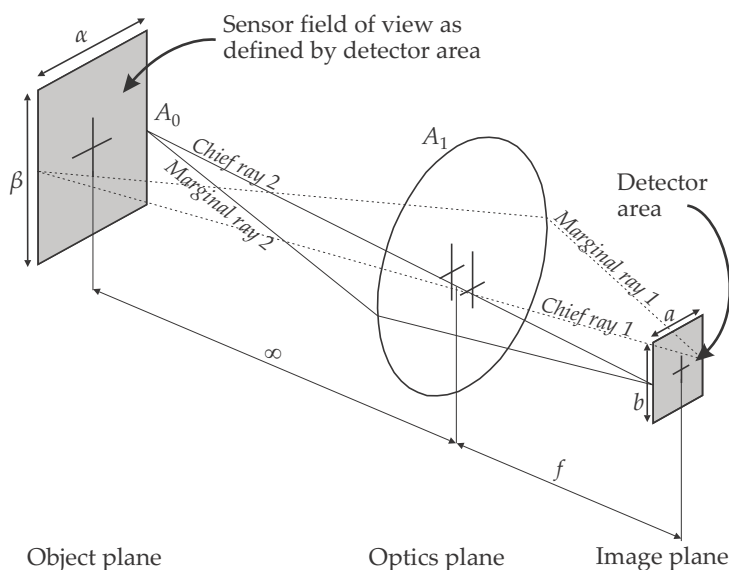


Figure 6.4 Sensor field of view defined by optical and detector parameters.

could be fixed (e.g., a metal plate cut to specific size and shape) or it could be variable (e.g., a variable iris for which the diameter of the opening can be adjusted). Figure 6.5(a) shows the apertures and stops in a real lens. In the thin-lens approximation, the entrance pupil falls in the plane of the thin lens, as shown in Figure 6.5(b).

Optical systems often suffer from unintended stray light when observing high-contrast scenes. The light could be reflected from optical surfaces or from the inside of the optomechanics. Stray light can be minimized by locating nonreflective stops or baffles so as to suppress the internal reflections.

A stop located in an image plane is called a field stop because this stop limits the field of view of the system. Some systems have more than one image plane (i.e., a telescope or microscope). By definition, these image planes are images of each other. The field stop can thus be located at any image plane and still have the same effect.

The pupil of an optical system is a plane that limits the amount of light flowing through the system. The lens diameters limit the amount of light passing through each lens, but because each lens limits the light in a different ‘plane’ of the system, the light control is not very precise and can lead to vignetting when some optical elements ‘shade’ other optical elements. Vignetting is the effect where one or more optical elements limit the amount of light nonuniformly over the image plane. The usual effect is

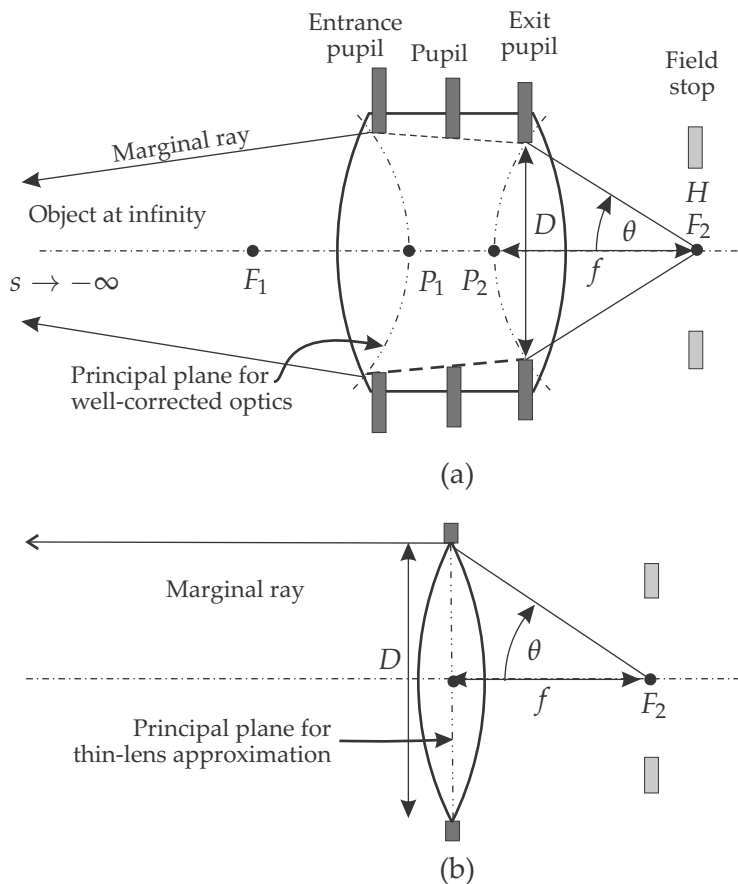


Figure 6.5 Stops, numerical aperture, and f -number: (a) complex (thick) lens and (b) thin-lens approximation.

that some (or all) edges of the image are darker than the center of the image. Severe vignetting may even cut off portions of the image completely, e.g., when the eye is not well aligned with the optical axis of a telescope.

One way to control vignetting is to accurately size the different optical elements appropriately. A better way to control vignetting is to make all of the elements slightly bigger than necessary and then to introduce a mechanical element (aperture stop) to limit the light in a single plane in a controlled manner. This approach also allows setting the aperture to different diameters, such as setting the aperture f -stop in a photographic camera.

The single element that limits the amount of light flow in the optical system is called the pupil. The pupil is normally a physical device inside the optics. The optical elements before the pupil image the pupil into

object space into the ‘entrance pupil.’ Likewise, the optical elements after the pupil image the pupil into image space into the exit pupil, i.e., seen from the image, this is the diameter of the optical system. The marginal ray in an optical system is the ray touching the edge of the pupil. In the thin-lens approximation, the entrance and exit pupil coincide in the thin lens and have the diameter of the thin lens, as shown in Figure 6.5.

The f -number ($f/\#$ or $F\#$) is a geometric construct comprising the diameter of the exit pupil and the focal length; it has no other meaning. The f -number is a convenient means to denote the amount of optical flux flowing through an optical system of a given focal length. The f -number of the optical system is defined as $f/\# = f/D$ for a system at infinite conjugates, where f is the focal length, and D is the diameter of the exit pupil. In photography the term f -stop is used to denote the size of a lens entrance pupil, in a progression of doubling areas for each successive f -stop. f -stop values are f -numbers, usually in a series of the form $f/1.4, f/2, f/2.8, f/4, f/5.6, f/8, \dots$. The optical systems encountered in this book are fixed f -number systems, and the f -number is only used to denote the ratio of focal length to diameter. Higher f -numbers result in lower irradiance on the focal plane. Systems with lower f -numbers (higher image-plane irradiance) become increasingly difficult to design with high optical performance, $f/1.2$ being a reasonable limit for good imaging quality. Generally speaking, a low- f -number lens is more complex and more expensive than a high- f -number lens.

The f -number is only defined for a system at infinite conjugates. For systems not at infinite conjugates, the ‘numerical aperture’ is used as a measure of the amount of flux flowing through the system at a given focal length. The numerical aperture is given by $NA = n \sin \theta$, where n is the index of refraction of the medium in which the lens is located, and θ is the angle of the marginal ray in image space, as shown in Figure 6.5. For lenses in air, $n = 1$. The numerical aperture is not constrained to infinite conjugates, and hence it can vary depending on the location of the image. The numerical aperture should not be estimated from the size of the aperture or lens diameter but from the inclination of the marginal ray (i.e., from the optical design).

At infinite conjugates the f -number and numerical aperture in image space are related by

$$NA = n \sin \theta = n \sin \left(\arctan \frac{D}{2f} \right) \approx n \frac{D}{2f}, \quad (6.4)$$

and hence

$$F\# = \frac{n}{2NA}. \quad (6.5)$$

This approximation is good for well-corrected optical systems. Why is a sine function and not a tangent function used in this equation? The reason is that the principal plane is not a mathematical plane, but in well-corrected systems, near-spherical, centered at F , as shown in Figure 6.5. The thin-lens approximation to f -number is therefore incorrect, and the numerical aperture is more meaningful.

In terms of flux collection, the numerical value of the f -number could be deceiving because it does not account for optical transmittance or obscuration in the optical pupil. A lesser-used construct is the t -number or t -stop, which is used to account for the transmittance or obscuration losses of an optical system. An optical system with such losses will have a t -number higher than the equivalent f -number for the same lens with no loss.

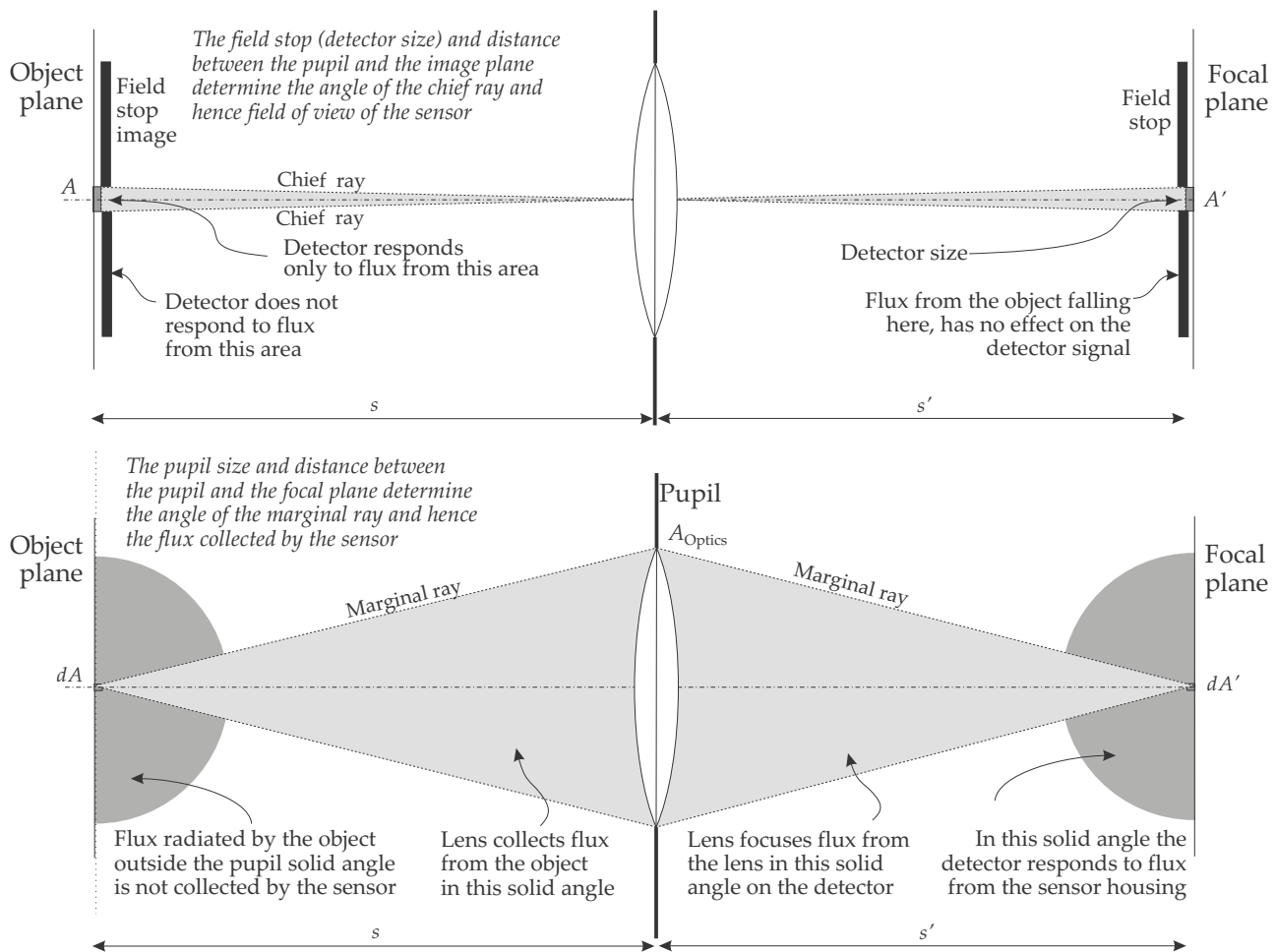
6.3.4 Optical sensor spatial angles

Following from the previous sections, it should be clear that there are two solid angles of concern in an optical sensor. These two solid angles are indirectly related to each other but have very little to do with each other. This section explores the field of view (FOV) solid angle and the image-flux-collecting solid angle, as depicted in Figure 6.6.

The flux-collecting property of an optical system is determined by the diameter of the pupil (the lens diameter) and is quantified by the numerical aperture or the f -number of the system. These two numbers are determined only by the marginal ray in the system. The bottom picture in Figure 6.6 shows the optical system and the marginal ray touching the pupil. It is evident from the bottom picture that the flux within the pupil is focused by the lens and collected in the detector on the focal plane. The irradiance in a point on the focal plane is determined only by the f -number/numerical aperture/marginal ray. The size of the detector has no direct effect on the flux-collecting performance of the optical system. The chief ray maps the field stop into object space, thereby determining the sensor's FOV. The size of the pupil has no direct effect on the FOV of the system.

Consider the top picture in Figure 6.6, the flux collected by the optical system: not all is collected within the field stop. The field stop blocks the flux outside of the desired field of view — this flux does not reach the focal plane. The sensor is only sensitive to flux within the field stop. The chief ray, determined by the size of the field stop and s' , therefore determines the sensor's FOV. The size of the pupil has no direct effect on the FOV of the system.

Figure 6.6 Optical sensor spatial angles.



There is a symmetry between the flux-collecting solid angle and the FOV: both are determined by an aperture or a stop and the value of s' . The common denominator is s' , but the commonality ends there. The FOV is determined by the field stop (size of the detector), whereas the flux-collecting solid angle is determined by the pupil or aperture stop. Section 6.8 describes a parameter called the system throughput, which considers the combined effect of the two solid angles.

6.3.5 Extended and point target objects

For the discussion in this section assume a uniform radiance over an object's area. If the object is smaller than the FOV, the object can be considered a point target in terms of the sensor FOV: the object's size has no effect on its signal in the FOV. If an object is larger than the FOV, it means that the full sensor FOV is filled by the object (called an extended target or object in terms of the pixel size) because the object area is bigger than the sensor FOV. The flux contributed by an extended object is defined by the size of the footprint of the sensor FOV on the object and not by the object size itself. See also Section 7.6.

6.3.6 Optical aberrations

Ideal, paraxial optical systems form perfect images; but practical optical systems do not because of optical aberrations.^{1-5,7} Optical aberrations are not only the result of poor workmanship or poor quality control but also a mathematically founded physical process. Elements with optical power (lenses or curved mirrors) refract or reflect optical rays toward a focal point; but a single point on the object may not always image to a single point in the focal plane.

Geometrical aberrations⁸ occur as a result of the geometry of the surfaces. There are huge families of physical aberrations, including several variations of spherical aberration, coma, astigmatism, and distortion. Figures 6.7 and 6.8 depict only the first-order aberrations. Spherical aberrations arise when rays distant from the optical axis focus in a different plane as do paraxial rays. As a result, there is a plane located at the 'circle of least confusion' that provides the smallest optical spot size. Chromatic aberration occurs as a result of the index of refraction in lenses not being constant for different colors. This means that blue light does not focus where green light focuses, and not where red light focuses. Astigmatism is an aberration when the optical power in the vertical and horizontal planes differ (e.g., cylindrical deviations from a spherical shape). Comatic aberration occurs when an off-axis spot is focused in different locations in the

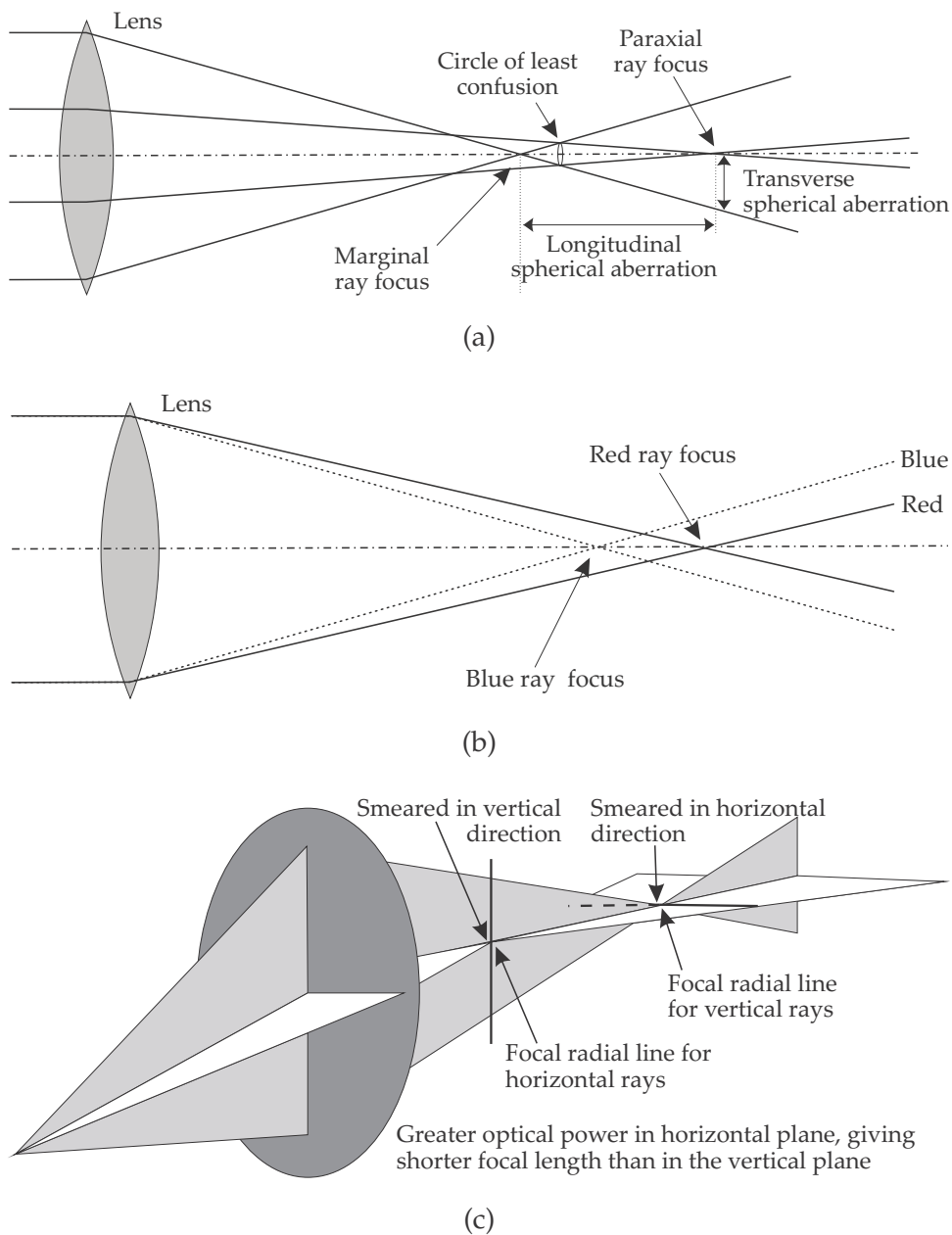
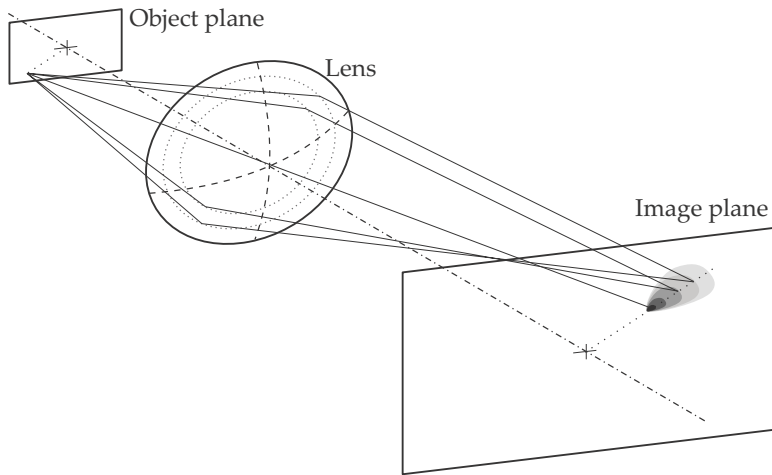
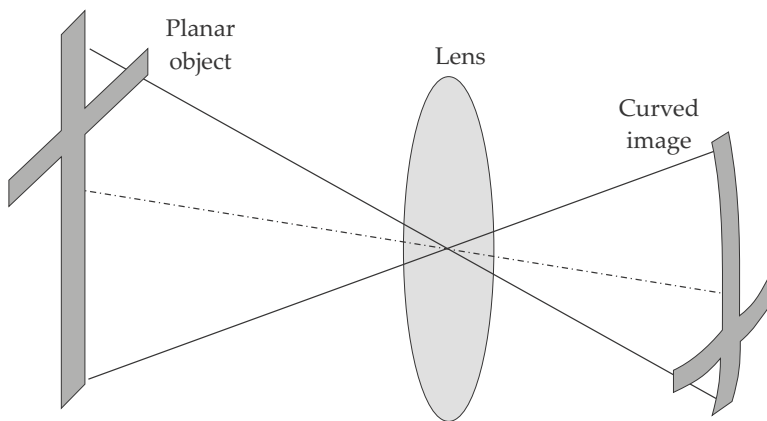


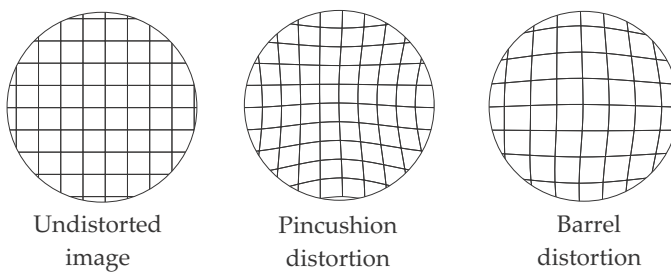
Figure 6.7 Optical aberrations: (a) first-order spherical aberration, (b) chromatic aberration, and (c) astigmatism.



(a)



(b)



(c)

Figure 6.8 Optical aberrations: (a) first-order coma, (b) field curvature, and (c) distortion.

focal plane by different parts in the lens. Coma gets its name from the comet-like shape of the spot. Field curvature is when the image of a plane in the object space is imaged to a curved surface. A focal surface with constant distance from the rear principle point (a simple lens) will have a spherical field curvature. Distortion arises when constant distances in the object plane are no longer constant in the image plane. The point spread function of off-axis aberrations (i.e., coma, astigmatism, and distortion) are not rotationally symmetric.

The main purpose of optical design is to create and employ degrees of freedom in the optical design so as to minimize the system's aberrations. These degrees of freedom include the number of optical elements, curvature of surfaces, thickness between surfaces, and index of refraction. The optical designer adjusts all of the degrees of freedom in the design to balance the optical aberrations against each other so as to cancel out, thereby providing the required optical performance. Optical design is a very sophisticated topic beyond the scope of this book.

6.3.7 Optical point spread function

The image sharpness of an optical system is described by the impulse response of a focused imaging system — this is also known as the point spread function (PSF) of the system. The mathematical Dirac delta impulse function can be conveniently modeled in optical terms by the infinitely small size of a distant star (assuming no atmospheric degradation), and the PSF is the image of the star. The PSF is a two-dimensional variable in the image plane.

Even a perfect optical system will have a non-Dirac delta impulse response, stemming from the physical diffraction of the light wave in the aperture. The impulse response of a clear circular aperture is called the Airy pattern,⁸ described mathematically using Bessel functions.

A practical optical system will have an impulse response comprising the diffraction effect, optical aberration effects, and aberrations due to manufacturing and optical element mounting. A lens design with negligible optical aberration is called a 'diffraction limited design' because diffraction is the main contributor to the optical PSF.

The optical PSF can be calculated from diffraction and aberration parameters. The optics' contribution to the PSF can be calculated by a technique called ray tracing, where the paths of large numbers of rays are calculated through the optical system. The PSF generally varies across the image: it is not a constant with change in field angle. From linear systems

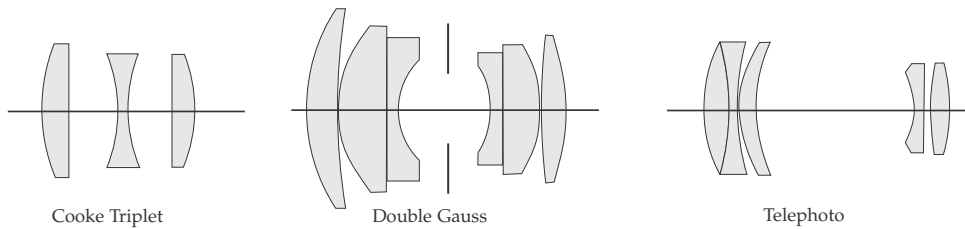


Figure 6.9 Commonly used refractive optical systems.

theory it can be shown that the image in the focal plane of an optical sensor will be the convolution of the PSF with the ‘ideal’ image in the absence of diffraction and aberrations.

The two-dimensional Fourier transform of the optical PSF yields the optical transfer function⁸ (OTF), a complex-value optical frequency response. The modulation transfer function (MTF) of the optical system is the magnitude of the OTF. The MTF of a system with aberrations is not always rotationally symmetric. MTF is commonly used for the specification, design, and testing of lenses.

6.3.8 Optical systems

Most optical systems consist of a number of optical elements. These elements are carefully shaped and located to achieve the required optical performance. Some typical lens designs are shown in Figure 6.9. Note the large differences between the designs in this figure: every lens design is unique and different from any other lens. Some lenses provide the required performance with only a few elements, whereas other lenses require a large number of elements to meet the requirements. Some low-cost applications, such as cameras in mobile phones, allow for large distortion in the optical design and correct the distortion by subsequent image processing.

Systems with long focal lengths may become very long and might not fit into the available space. A very convenient solution is to ‘fold’ the system with mirrors, resulting in a shorter and more compact system. Figure 6.10 shows a Cassegrain folded optical system. This approach yields shorter and more-compact solutions than the equivalent unfolded design. Another folded optics system is the reflective Gregorian afocal telescope. Afocal means that the optical system does not have a focal length; it does not focus an object to an image plane. The Gregorian telescope has folded optics, but it is not necessarily shorter than the refractive equivalent. Reflective optics may yield shorter optical systems but carry

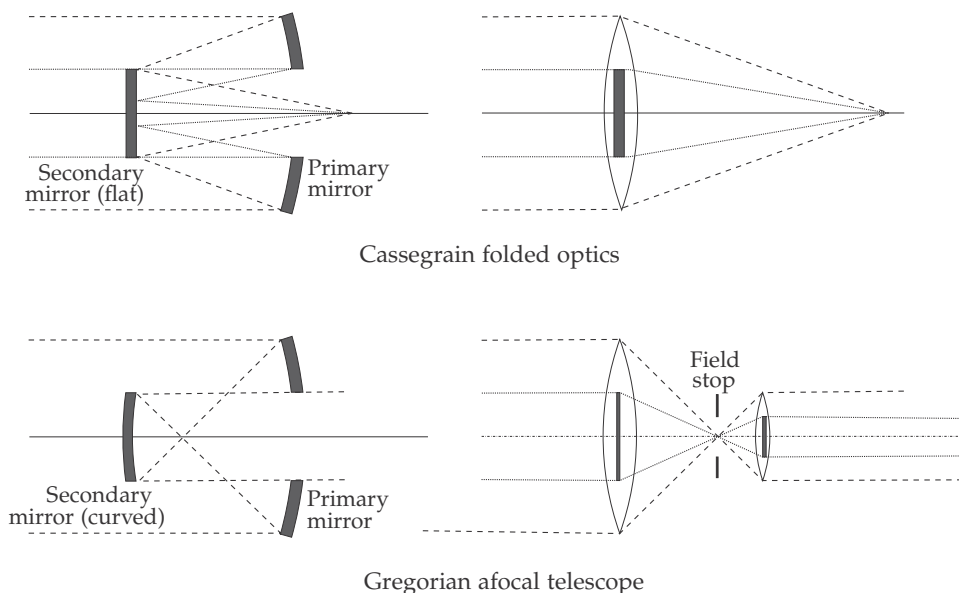


Figure 6.10 Reflective optics.

several disadvantages, including central obscuration, manufacturing, and packaging difficulties.

6.3.9 Aspheric lenses

The optical power in an optical surface stems from the curvature of the optical surface. In principle, any curved surface can produce optical power, as is evidenced by studying common, household glass items. Most optical elements were traditionally manufactured with spherical profiles because such profiles are easily formed by conventional lapping and polishing techniques. There is also a class of surfaces called aspherical surfaces with useful optical properties.

The profile of an aspheric lens is neither spherical, cylindrical, nor flat. During optical design the surface of an aspherical element can be specified such that it provides optical power, but at the same time minimizes or corrects optical aberrations caused by other elements. In particular, the aspherical surface reduces or eliminates spherical aberration. This means that fewer optical elements are required, and the lens assembly is simpler and lighter. Aspherical elements are more difficult to manufacture than spherical optics and are thus used where the total system solution provides benefits such as weight, complexity, or cost.

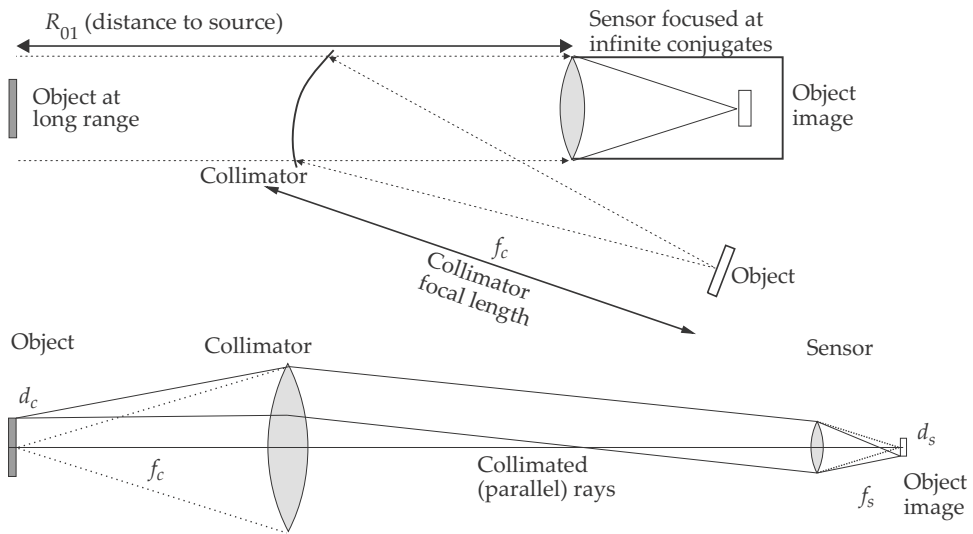


Figure 6.11 Laboratory collimator and long range configurations.

6.3.10 Radiometry of a collimator

Many sensors are focused at infinite conjugates, hence there is need for an instrument that can create infinite object images within the limited confines of the laboratory. A collimator is simply an element with optical power (lens or mirror), focusing an object at infinity. Because the image is located at infinity, the optical rays are parallel to the optical axis (collimation).

The setup can be depicted as shown in the top picture in Figure 6.11. In terms of the flux transfer defined in Equation (2.31), the object range R_{01} must be replaced by the collimator focal length. The sensor is focused at infinity and thus can image the collimated object image, in focus, on the sensor focal plane. The image diameter in the sensor focal plane is given by $d_s = d_c f_s / f_c$, where f_s is the sensor optics focal length, f_c is the collimator focal length, and d_c is the object diameter.

The collimator provides an ideal means to perform laboratory work at infinite conjugate sensors, but there are practical considerations complicating such work. These considerations include optical axis alignment, lateral placement in the beam, and beam vignetting. If the collimator and sensor optical axes are not parallel, the object image will be displaced from the center of the sensor image. If the sensor is displaced laterally, part of the flux in the beam will be lost.

One common use of a collimator is to simulate a source at long range in the laboratory. Figure 6.12 illustrates the vignetting effect that occurs

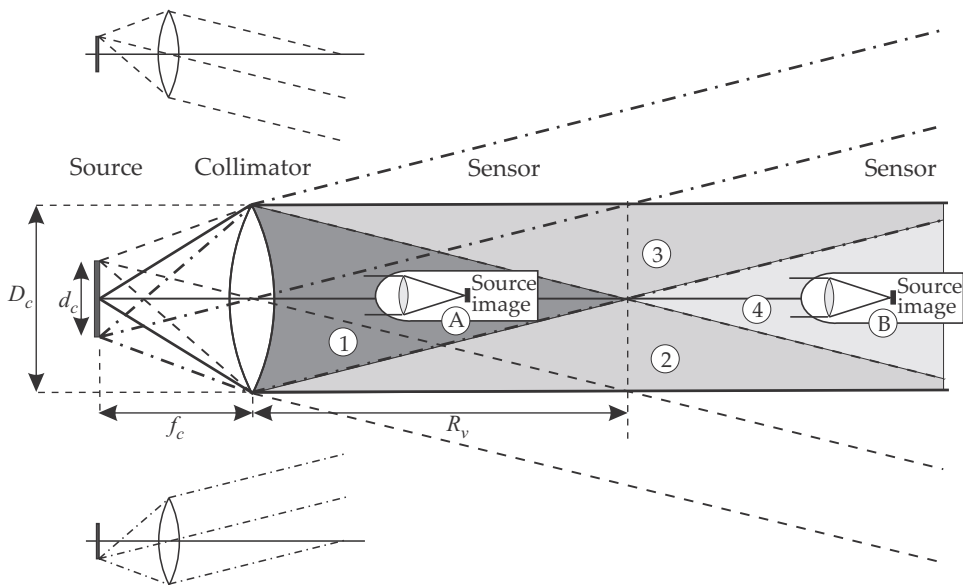


Figure 6.12 Collimator beam vignetting.

for a large source and small collimator focal lengths. Flux emanating from the center of the source is collimated to parallel rays (solid lines in Figure 6.12). Flux emanating from the outer edges of the object is collimated to parallel rays at an angle to the optical axis (see the two smaller inserts in Figure 6.12).

From the geometry, it is evident that the only part of the beam providing flux from the full source area is the conical-shaped zone 1. Zone 4 does not contain any flux from the outer rim of the source, whereas zones 2 and 3 vignette either the bottom or top part of the object area. This simple geometrical analysis does not take into account the sensor's FOV and entrance aperture diameter, both of which place even further restrictions on the sensor placement within zone 1. If it is important to obtain all of the flux emanating from the source (i.e., when calibrating a sensor), it is necessary to locate the whole of the sensor aperture in zone 1. The sensor located at position A in the beam has no vignetting. The sensor located at position B in the beam will only receive flux from the central part of the source.

Using similar triangles, it follows that $R_v/D_c = f_c/d_c$, hence $R_v = D_c f_c/d_c$. The depth of nonvignetted zone 1 can be increased by using a larger collimator (larger D_c and f_c) or a smaller source (smaller d_c). A smaller source area might not be feasible because the source area is normally selected to control the total amount of flux from the source.

6.4 Spectral Filters

An optical filter is an element that transmits light selectively. Several physical processes can be used to effect the filtering action. The two most-common techniques used are interference filters and absorption filters. In the case of absorption filters, the bulk of the filter material absorbs energy. The interference filter consists of many thin film dielectric layers acting as resonators at optical frequencies. The filter can effectively pass selected frequencies but reject other frequencies. The class of neutral density filters provides a constant spectral transmittance over all wavelengths.

Filters are designed to the following generic spectral requirements: the *passband*, where the filter should transmit light with the specified minimum transmittance, and the *stopband*, where the filter should suppress light to the specified suppression. A typical transmittance specification might be that the filter has a transmittance exceeding 0.7 in the spectral range 0.7–0.86 μm , with a transmittance of less than 0.001 outside the spectral range 0.65–0.9 μm , over the spectral range of a silicon detector defined as 0.3–1.2 μm . In the absence of measured filter transmittance data, the spectral response of a filter can be approximated by Equation (D.4), with potential shapes shown in Figure D.1(b).

A spectral filter is a physical object, subject to Planck's law and Equation (2.2). At wavelengths where the filter transmittance is less than unity, the filter reflects ambient flux and/or radiates thermally generated flux. This leads to the phenomenon where the sensor observes the filter as a source of energy (in the stopbands). In order to reduce the filter radiation, the filter is sometimes cooled down, even down to cryogenic temperatures.

6.5 A Simple Sensor Model

This section describes a minimum detail model of a sensor in terms of simple optics, a detector, and a spectral filter. The very simple sensor model, developed here from first principles, is further developed in later chapters.

The sensor is regarded as an imaging system, projecting an image from a distant object onto the focal plane of the lens. The basic (thin lens) configuration is shown in Figure 6.13. This combination of lens and detector results in a sensor that is spatially sensitive over a small angle, called the field angle or the FOV. In other words, the sensor is insensitive to radiation coming from sources outside its FOV. It is almost the inverse of a flashlight, with the lamp filament being replaced with a detector. Compare

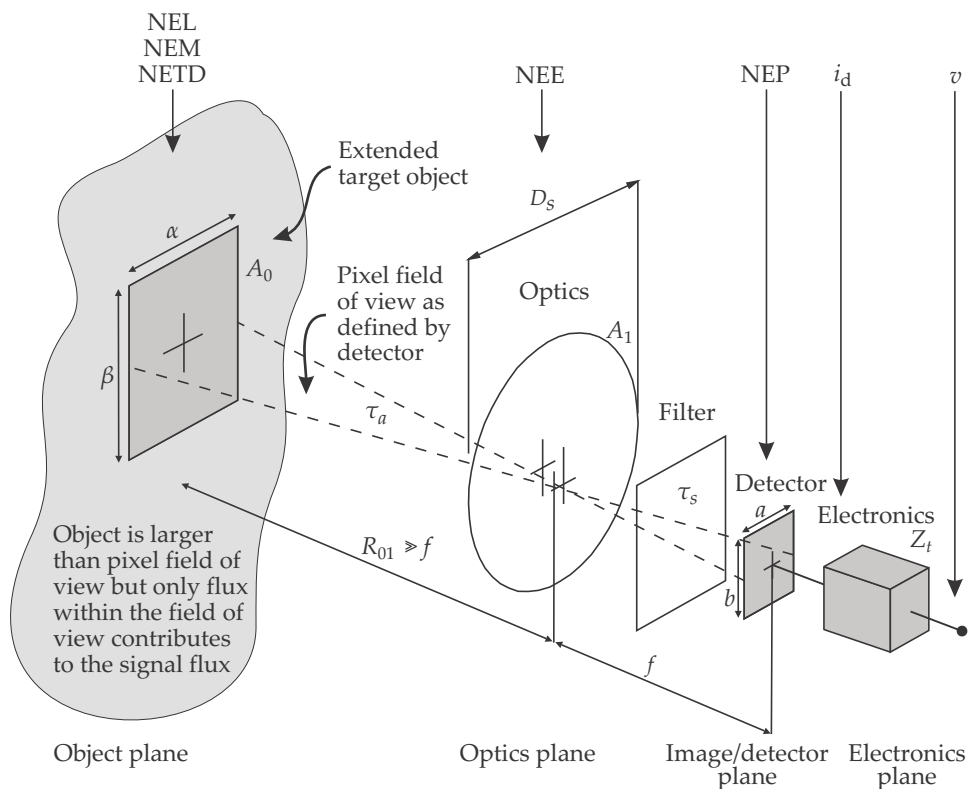


Figure 6.13 System parameters used in the simple sensor model.

Figure 6.13 with Figure 6.6, specifically with respect to the role of the optics in this sensor.

Most sensor systems operate at or near infinite conjugates; for this model it is assumed that the sensor is focused at infinity $s \rightarrow -\infty$, and hence $s' \rightarrow F_2$. Accurate mathematics require that $\tan \alpha = a/f$, but for small angles $\tan \alpha \approx \alpha$. The respective FOV angles for *small angles* are then given by the *chief ray* in terms of the detector size and optics focal length by the following simplified equations:

$$\alpha = \frac{a}{f}, \quad (6.6)$$

$$\beta = \frac{b}{f}, \text{ and} \quad (6.7)$$

$$\begin{aligned} \omega &= \alpha\beta \\ &= \frac{ab}{f^2}. \end{aligned} \quad (6.8)$$

These equations state that a lens with focal length f , in [m], with

a detector with dimensions a and b , in [m], in the focal plane has FOV angles α and β in [rad]. The sensor pixel FOV has a solid angle FOV ω in [sr]. For focus at infinite conjugates the optics diameter and clear aperture area are described in terms of the f -number and detector size by

$$D_s = \frac{f}{F_{\#}}, \text{ and} \quad (6.9)$$

$$A_s = \frac{P\pi D_s^2}{4} \quad (6.10)$$

$$= \frac{P\pi f^2}{4F_{\#}^2} \quad (6.11)$$

$$= \frac{P\pi ab}{4\omega F_{\#}^2}, \quad (6.12)$$

where $F_{\#}$ is the f -number of the optics, D_s is the optics aperture diameter in [m], A_s is the clear aperture area in [m²], and the factor P is used to account for diverse loss effects such as vignetting or central obscuration.

6.6 Sensor Signal Calculations

Extending on the optical sensor model developed in Section 6.5, this section develops a radiometric model of a simple source–medium–sensor system to calculate the signal at various locations in the system. In terms of Equation (2.33) the source is modeled as a Planck radiator with a spectral emissivity $\epsilon_{0\lambda}$. The elemental area dA_1 is now the entrance pupil of the optical system, focusing flux onto a detector. It is common practice to locate an optical filter with transmittance $\tau_{s\lambda}$, near the pupil A_1 in order to selectively control the flux flowing through the detector. An amount $\alpha_{s\lambda}$ is absorbed in the receiver area (in most cases this value $\alpha_{s\lambda}$ is already accounted for in the detector responsivity). The detector converts the flux flowing through A_1 to an electronic signal. This signal is amplified and can be used as a measure of the optical flux through the sensor's entrance pupil.

6.6.1 Detector signal

The source and receiver areas are often small, relative to the distances involved. Then, $\int dA_1 \rightarrow A_1$. Applying Equations (2.33) and (3.25), the spectral flux in the entrance pupil of a sensor can be written as

$$d^2\Phi_{\lambda} = \frac{\epsilon_{0\lambda} L_{0\lambda} dA_0 \cos\theta_0 A_1 \cos\theta_1 \tau_{a\lambda} \tau_{s\lambda} \alpha_{s\lambda}}{R_{01}^2} \quad (6.13)$$

for the flux in a narrow spectral band and small source area. The $\alpha_{s\lambda}$ factor is the flux absorptance (emissivity by Kirchhoff's law) in the sensor

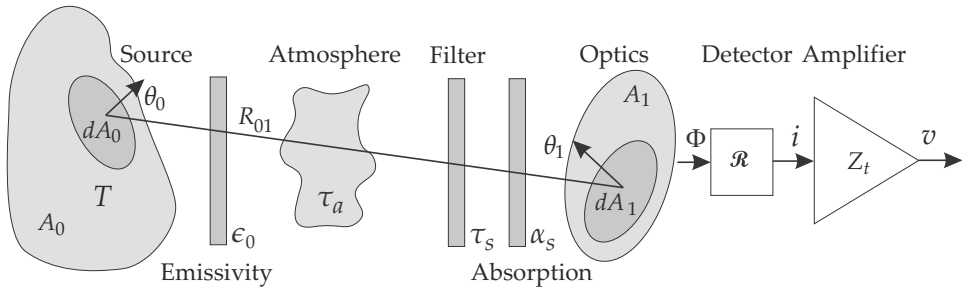


Figure 6.14 Simple radiometric model of a source, medium, and sensor.

as explained in Figure 3.5. It is henceforth assumed that the sensor is aligned toward the object such that $\theta_1 = 0$ so that the $\cos \theta_1$ becomes unity. Any real sensor consists of several optical elements, which may include a window, reflective optical elements (mirrors), refractive optical elements (lenses), and optical filters. The spectral transmittance values of all of these filters are assumed to be contained in $\tau_{s\lambda}$.

As shown in Figure 6.14, the flux on the entrance aperture is converted to an electrical signal in the detector and amplified by the amplifier. Assume that a fraction k of the flux falling on the entrance aperture reaches the detector. The value of k is normally less than unity because of losses in the optics, vignetting, or other attenuating mechanisms in the sensor. The detector spectral responsivity may be written as a normalized spectral responsivity multiplied by a peak responsivity, $\mathcal{R}_\lambda = \hat{\mathcal{R}} \tilde{\mathcal{R}}_\lambda$ (normally with units of [A/W] or [V/W]). The absorptance variable $\alpha_{s\lambda}$ in Equation (6.13) is contained in the detector spectral response $\tilde{\mathcal{R}}_\lambda$. The detector signal is normally fed into an integrator or preamplifier to capture and amplify the signal for further processing. Assuming an ideal linear preamplifier and an electronic bandwidth that would not limit the signal, the preamplifier output for a monochromatic source at wavelength λ is given by

$$d^2v_\lambda = \frac{k \mathcal{R} Z_t \epsilon_{0\lambda} L_{0\lambda} dA_0 \cos \theta_0 A_1 \tau_{a\lambda} \tau_{s\lambda} d\lambda}{R_{01}^2}, \quad (6.14)$$

where Z_t is the preamplifier gain in [V/A]. Define the spectral *sensor response* as

$$\mathcal{S}_\lambda = \tau_{s\lambda} \tilde{\mathcal{R}}_\lambda, \quad (6.15)$$

regroup the variables, and integrate over all wavelengths, to obtain the signal voltage as

$$dv_S = \frac{k \hat{\mathcal{R}} Z_t dA_0 \cos \theta_0 A_1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda, \quad (6.16)$$

where v_S is the detector voltage [V] for flux in the band defined by S , and the elemental source area dA_0 has been retained. Equation (6.16) can be integrated over the source area to obtain the full source signature:

$$v_S = k \hat{\mathcal{R}} Z_t A_1 \int_{A_0} \left(\frac{1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} S_\lambda d\lambda \right) dA_0 \cos \theta_0. \quad (6.17)$$

Note that Equation (6.17) can be calculated in either the radiant (watts) or photon (photons/second) domains. In the radiant domain, use Equations (3.1) and (5.5). In the photon domain, use Equations (3.3) and (5.4).

It is convenient to define an irradiance, called the *apparent irradiance*, or sensor inband irradiance, which is the irradiance from a given source to a given sensor observed through a given medium. The apparent sensor inband irradiance is defined as

$$E_S = \frac{\Phi_\lambda}{A_1} \quad (6.18)$$

$$= \frac{k dA_0 \cos \theta_0}{R_{01}^2} \int_{\lambda_1}^{\lambda_2} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} S_\lambda d\lambda. \quad (6.19)$$

From Equation (6.19) it can be seen that the source radiance is spectrally weighted with the system response and the medium transmittance, hence the name *apparent irradiance*: ‘as observed by the sensor.’ Equations (6.15), (6.16), and (6.19) form the basis of the system model. These equations calculate the expected signal for any object at any distance. Because the spectral integrals contain factors of the source, the atmosphere, and the sensor, it is necessary to calculate the integral for any new set of sensor/object data or change in atmosphere or distance.

6.6.2 Source area variations

The previous section assumed a small and uniform source with area dA_0 . By integrating Equation (6.16) across the source area A_0 , the total detector signal due to an arbitrary source can be determined as

$$v_S = k \hat{\mathcal{R}} Z_t A_1 \int_{\text{source}} \left(\int_{\lambda_1}^{\lambda_2} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} S_\lambda d\lambda \right) \frac{d(A_0 \cos \theta_0)}{R_{01}^2}, \quad (6.20)$$

where the spectral integral is calculated first for each elemental area dA_0 , and then over the full area A_0 . Note that $L_{0\lambda}$ may be a function of dA_0 , i.e., the source radiance may vary over the surface of the source.

In the case where the source size is small compared with the distance between the source and the detector, R_{01} reduces to the distance between

the source and the detector. If in addition, the source has uniform radiance over its total area, and the source solid angle is less than the sensor field of view, the spatial integral over the source can be separated from the spectral integral:

$$v_S = k \widehat{\mathcal{R}} Z_t A_1 \int_{\lambda_1}^{\lambda_2} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda \int_{\text{source}} \frac{d(A_0 \cos \theta_0)}{R_{01}^2}. \quad (6.21)$$

The source spatial integral is actually the source solid angle

$$\Omega_s = \int_{\text{source}} \frac{d(A_0 \cos \theta_0)}{R_{01}^2}, \quad (6.22)$$

and the detector signal can be written as

$$v_S = \Omega_s k \widehat{\mathcal{R}} Z_t A_1 \int_{\lambda_1}^{\lambda_2} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda. \quad (6.23)$$

Equation (6.23) can be applied for relatively small, uniform sources where the source solid angle, as subtended at the sensor, is less than the sensor FOV. If the sensor FOV is smaller than the source solid angle, the source solid angle in Equation (6.23) is replaced with the sensor solid angle, as explained in Section 7.6.

6.6.3 Complex sources

The mathematical model described thus far only considers one source element. A complex source may contain a number of different radiators with highly different characteristics. One portion may be very hot and another portion may be much cooler. For such a source, the total irradiance, within a single FOV, can be found by summing the contributions from all N elements as follows:

$$v_S = k \widehat{\mathcal{R}} Z_t A_1 \sum_{i=0}^N \left(\frac{A_{0i} \cos \theta_{0i}}{R_{01i}^2} \int_{\lambda_1}^{\lambda_2} \epsilon_{0\lambda i} L_{0\lambda i} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda \right). \quad (6.24)$$

6.7 Signal Noise Reference Planes

In a real, physical system, the noise contributions are typically created in the detector and the first-stage electronic amplifier. It is sometimes required to express the noise or signal *referred* to different locations in the system. For example, the sensor noise can be expressed as an irradiance on the entrance pupil of the sensor or even to the source/object plane. The referred noise or signal levels have the same effect as in the detector

but are appropriately scaled. The different locations are called ‘planes’ following the practice in optical design. As shown in Figure 6.13 there are four planes of importance: the *object plane* at the source object, the *entrance pupil* or *optics plane* at the entrance pupil of the optics, the *image plane* or *detector plane*, and the *electronics plane*.

The mechanism to convert noise is based on the standard radiometry methodology. The same transformations that are used for optical flux are used exactly the same way for noise transformations. In particular, note the following relationships:

$$E = \frac{LA_0 \cos \theta_0}{R^2} = L\omega = \frac{M\omega}{\pi}, \quad (6.25)$$

which are extended to the equivalent noise representations

$$NEE = \frac{NEL A_0 \cos \theta_0}{R^2} = NEL \omega = \frac{NEM \omega}{\pi}. \quad (6.26)$$

Note that in the above equation, the projected solid angle should be used because the thermal camera senses Lambertian sources. However, for the small pixel fields of view generally used, the projected and geometrical solid angles are numerically equal.

The conventions followed in this book are as follows (the same conventions apply to both noise and signal):

1. Spectral noise equivalent power (NEP_λ) is the optical signal power spectral density [$W/\mu m$] required on the detector to yield a signal equal to the noise in the sensor. NEP_λ is always measured in the detector plane and is expressed as flux in the image plane. See also Section 7.1.3.
2. Spectral noise equivalent irradiance (NEE_λ) is the optical signal irradiance required in the entrance pupil to yield a signal equal to the noise in the sensor. NEE_λ is always measured in the entrance pupil of the sensor (but excluding the sensor filter) and is expressed in [$W/(m^2 \cdot \mu m)$] incident on the sensor. NEE_λ can be conceptually defined as

$$NEE_\lambda = \frac{NEP_\lambda}{A_1 \tau_s} = \frac{\sqrt{\Delta f A_d}}{D_\lambda^* A_1 \tau_s}, \quad (6.27)$$

where A_1 is the sensor entrance pupil area, Δf is the noise equivalent bandwidth, A_d is the area of the detector, D_λ^* is the spectral specific detectivity, and τ_s is the sensor system response (filter transmittance and other losses between the entrance pupil and detector). If $\tau_s = 1$, it means that all of the power incident on the entrance pupil reaches the detector. See also Section 7.1.3.

3. Spectral noise equivalent radiance (NEL_λ) is the radiance required at the source in the object plane to yield a signal equal to the noise in the sensor. Units are $[W/(m^2 \cdot sr \cdot \mu m)]$. NEL_λ is always measured in the object plane.

NEL_λ is defined only for extended targets (filling the complete FOV) and can be derived from the NEP by using Equation (2.31):

$$NEL_\lambda = \frac{NEP_\lambda R_{01}^2}{A_0 A_1 \tau_s \tau_a}, \quad (6.28)$$

where A_0 is the pixel footprint in the object space, R_{01} is the distance between the sensor optics and the object plane, and τ_a is the atmospheric transmittance. See also Section 7.1.3.

4. Spectral noise equivalent exitance (NEM_λ) is the exitance required at the source to yield a signal equal to the noise in the sensor. NEM_λ has units $[W/(m^2 \cdot \mu m)]$ and is always measured in the object plane. The NEM_λ can be derived from the NEL_λ by $NEM_\lambda = \pi NEL_\lambda$ [by Equation (2.7)]. See also Section 7.1.3.
5. Spectral noise equivalent temperature difference ($NETD_\lambda$) is the change in source temperature [K] required to yield a signal equal to the noise in the sensor. $NETD_\lambda$ is always measured in the object plane and is expressed in temperature difference at the object. $NETD_\lambda$ is derived from the NEL_λ by noting that for small temperature variations

$$\frac{NEL}{NETD} = \frac{dL}{dT}, \quad (6.29)$$

and hence

$$NETD_\lambda = \frac{NEP_\lambda R_{01}^2}{A_0 A_1 \tau_s \tau_a \frac{dL}{dT}}, \quad (6.30)$$

where $NETD_\lambda$ has units of [K], and dL/dT is Planck's law temperature derivative [see Equation (3.4)]. See also Sections 7.1.3 and 9.6.

Note that the sensor system response τ_s , atmospheric transmittance τ_a and NEP used above are spectrally varying variables. The above equations are therefore only valid at a particular wavelength, not over a wide wavelength band. Wideband spectral quantities are discussed in more detail in Section 2.6.5.

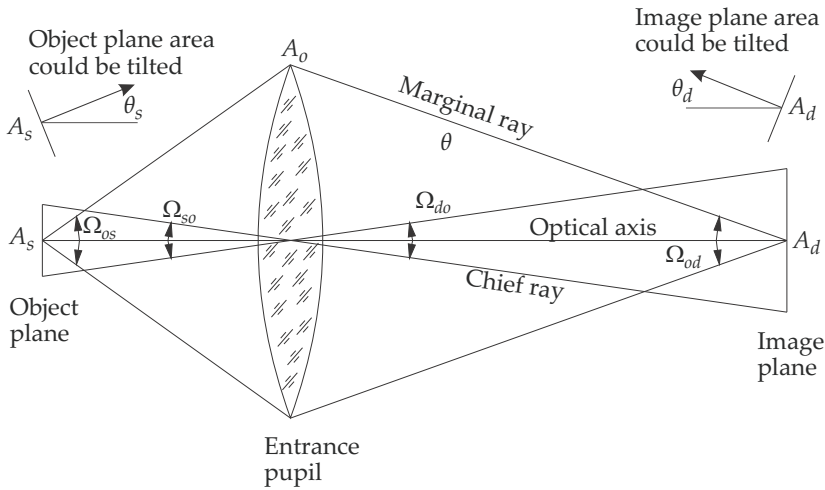


Figure 6.15 Radiance in imaging systems.

6.8 Sensor Optical Throughput

Equation (2.31) forms the basis of all radiometric calculations; the flux flowing from a source to a receiving area can be written as

$$\begin{aligned}
 d^2\Phi &= \frac{L dA_0 \cos \theta_0 dA_1 \cos \theta_1}{R_{01}^2}, \\
 &= L d\Omega_0 dA_1 \cos \theta_1, \\
 &= L dA_0 \cos \theta_0 d\Omega_1,
 \end{aligned} \tag{6.31}$$

which implies that the total flux is proportional to the (spatially invariant) radiance times the product of the source area and the receiver solid angle, or vice versa. It follows that the source could have been on either side; *the total flux transfer depends on the geometrical relationship between the two areas, irrespective of which is the source.*

For the simplified lossless imaging system shown in Figure 6.15, the argument above can be extended further. It can be shown that the total flux flowing through the source and detector is given by

$$\begin{aligned}
 d^2\Phi &= L d\Omega_{so} dA_o \cos \theta_o, \\
 &= L d\Omega_{os} dA_s \cos \theta_s, \\
 &= L d\Omega_{od} dA_d \cos \theta_d, \text{ and} \\
 &= L d\Omega_{do} dA_o \cos \theta_o.
 \end{aligned}$$

In all of the above equations it is evident that the total flux flowing through the system is proportional to the product of a solid angle and the

appropriate projected area. This product of solid angle and area is called the throughput of the system. System *basic throughput*, or *étendue*, is an indication of the total flux that can pass through the system. It depends on the FOV and the aperture of the system. It is defined by

$$T = n^2 A \Omega, \quad (6.32)$$

where n is the refractive index at the location where the solid angle Ω is defined, and A is the area of the aperture at the location where the solid angle is defined. In terms of the optical system of Figure 6.15, the étendue is given by $n_o^2 A_o \Omega_s$. The units of throughput are $[\text{sr} \cdot \text{m}^2]$.

In its most-basic form, the power flowing through the source and receiver areas is given by the product of the basic radiance and the system throughput,

$$\begin{aligned} \Phi &= \left(\frac{L}{n^2} \right) n^2 A \Omega, \\ &= \frac{L A_0 A_1}{d^2}. \end{aligned}$$

Optical designers define a quantity, called the Lagrange invariant, that is invariant in any given optical system. The Lagrange invariant is proportional to the square root of the throughput.

The concept of throughput underlines a very important principle: for a fixed optics f -number and a fixed image height (e.g., detector size), the product of the look angle (solid angle) and receiving aperture (area), and hence the flux through the system, is constant. Under this condition, any increase in FOV requires a decrease in the optics aperture and vice versa:

$$T = n^2 A_{\text{optics}} \Omega_{\text{fov}} = n^2 \left(\frac{\pi f^2}{4F_{\#}^2} \right) \left(\frac{ab}{f^2} \right) = \frac{n^2 \pi ab}{4F_{\#}^2} = \text{constant}. \quad (6.33)$$

Consider two lenses for use with a staring array detector with a diagonal size of 13 mm. The two diagonal field angles required are 5 deg and 25 deg. Assume that both lenses have apertures of $f/1.8$. First-order calculations yield the lens specifications shown in Table 6.1. It is clear that the wide-angle lens has a smaller diameter, and the narrow angle lens has a much bigger diameter. This ‘unfortunate’ fact is due to the limitation imposed by the throughput of the system (fixed image height and f -number).

When a sensor with given throughput (fixed detector size and f -number) views an extended target object, the power on the detector stays the same, irrespective of the FOV. Hence, the sensor gain calibration is identical for all fields of view (for the given f -number).

Table 6.1 Example lens design demonstrating optical throughput.

Field angle	Focal length	Aperture diameter
5	149	82
25	29.8	16
deg	mm	mm

The throughput limitation is only imposed on optical systems that attempt to obtain some measure of image quality. If the optical system is not designed for reasonable image quality, as in typical energy-concentrating optics, this throughput limitation does not apply. Energy-concentrating systems can therefore have large apertures and large fields of view.

Bibliography

- [1] Shannon, R. E., *The Art and Science of Optical Design*, Cambridge University Press, Cambridge, UK (1997).
- [2] Fischer, R. E., Tadic-Galeb, B., and Yoder, P. R., *Optical System Design*, McGraw-Hill, New York (2008) [doi: 10.1036/0071472487].
- [3] Smith, W. J., *Modern Lens Design*, 2nd Ed., McGraw-Hill Professional (2004).
- [4] Smith, W., *Modern Optical Engineering*, 4th Ed., SPIE Press, Bellingham, WA (2007).
- [5] Kingslake, R. and Johnson, R. B., *Lens Design Fundamentals*, 2nd Ed., SPIE Press, Bellingham, WA (2010) [doi: 10.1016/B978-0-12-374301-5.00003-6].
- [6] Walker, B., *Optical Engineering Fundamentals*, 2nd Ed., SPIE Press, Bellingham, WA (2009).
- [7] Schmidt, J. D., *Numerical Simulation of Optical Wave Propagation: With examples in MATLAB*, SPIE Press, Bellingham, WA (2010).
- [8] Hecht, E., *Optics*, 4th Ed., Addison Wesley, Boston, MA (2002).

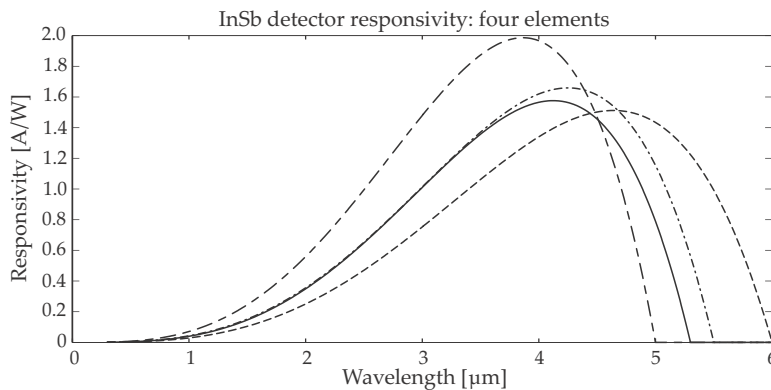
Problems

- 6.1 An optical system with an f -number of $f/5$ employs a single-element detector with size $1 \times 1 \text{ mm}^2$. The focal length is 250 mm.

- 6.1.1 Determine the instantaneous solid FOV of the sensor (in radians). [2]
- 6.1.2 Determine the geometric solid angle subtended by the lens as seen from the detector. [2]
- 6.1.3 Determine the projected solid angle that the barrel (the walls of the tube holding the lens and detector) of the sensor subtends as seen from the detector. [2]
- 6.1.4 Determine the throughput of the sensor. [2]
- 6.2 Derive a mathematical description of the detector current for a simple system comprising a small detector, an optical filter, an atmospheric medium, and a small thermal source. The detector and source are separated by more than 100 times the diameter of the detector. This system has no other optics, collimators, or components. Start with the simple graphic for flux transfer between two small surfaces. Apply the Golden Rules when developing the answer. [6]
- 6.3 Use Equations (6.1), (6.2), and (6.3) to calculate the image height, the magnification, and distance from the back principal plane (s') for a lens with focal length 100 mm, an object of size 10 mm, at distances of 200, 400, 1000, and 10000 mm. Show the results in tabular form. [4]
- 6.4 Use a 1:1 scale geometrical diagram to plot the chief ray and marginal ray to calculate the image location and image height for a lens focal length of 50 mm, an object height of 10 mm, with the object located at -100 , -200 , and $-\infty$ mm. The object located at infinity does not have to be to scale because it would require an infinite-sized paper.

Confirm the results obtained with the geometrical plot by using the equations above to calculate the image height and image location. [3]
- 6.5 A sensor has a lens with focal length of 200 mm, f -number of 3, and a detector size of $10 \times 10 \text{ mm}^2$. Draw a diagram of the sensor to scale, showing the chief ray and marginal ray of the system. Calculate the numerical aperture, the FOV solid angle of the sensor, and the light-collecting solid angle (defined by the numerical aperture) at the detector. Confirm the calculation by comparing the results with the angles measured or calculated in the scale drawing. [5]
- 6.6 A simple sensor has four pixels, but each of the pixels has a slightly different responsivity, as shown below. This means that

if the sensor views a uniform background, the pixels in the image have different values - this is called detector nonuniformity. The objective with this task is to design a means to perform non-uniformity correction (NUC). The spectral responsivity is shown here for interest only, it is not required in the execution of the task.



Each detector element size is $1 \text{ mm} \times 1 \text{ mm}$. The detector is located 200 mm from a blackbody. The blackbody aperture (size) is $1 \times 10^{-6} \text{ m}^2$. Two detector current measurements are performed at different blackbody temperatures. The 0–6- μm inband powers used in the measurements are as follows:

Temperature	700 K	900 K	1000 K
inband power Φ	0.0560 μW	0.2016 μW	0.3331 μW

The measured currents in the four detector elements are:

	i_{700}	i_{900}	i_{1000}
Temperature	700 K	900 K	1000 K
No. D1	0.0642	0.1954	0.2991
No. D2	0.0655	0.2324	0.3763
No. D3	0.0571	0.1903	0.3012
No. D4	0.0636	0.2072	0.3253
-	μA	μA	μA

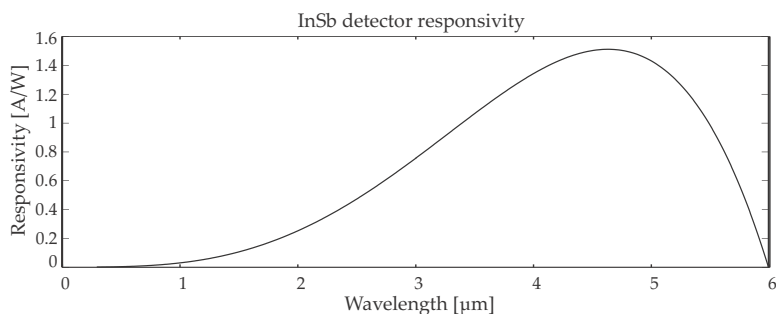
- 6.6.1 Plot the detector current versus input power for the four elements. Observe that there are four different response lines. [4]
- 6.6.2 A vertical line on the graph represents the optical power from a uniform source. Predict what the detector current values will be for an optical power of 0.25 μW . [2]
- 6.6.3 Comment on how the signals from the four elements can be compensated (by external electronics) such that the sensor output sig-

nal will be the same for all four detectors, irrespective of the input power. [2]

- 6.6.4 Using a block diagram only, do a conceptual design of a circuit that will perform a nonuniformity correction (NUC) for each of the four elements in this sensor. [4]

- 6.7 A lens with focal length 120 mm and diameter of 50 mm is located in front of a square detector, with the detector in the focal plane of the lens. The lens barrel totally encloses the rear of the sensor (from the lens, backwards, all around the detector). The square InSb detector has an area of 0.1 cm^2 . The sensor is placed in the center of a cubic oven with wall dimensions of 3 m, pointing to the center of one of the walls. The oven walls are heated to 1300 K.

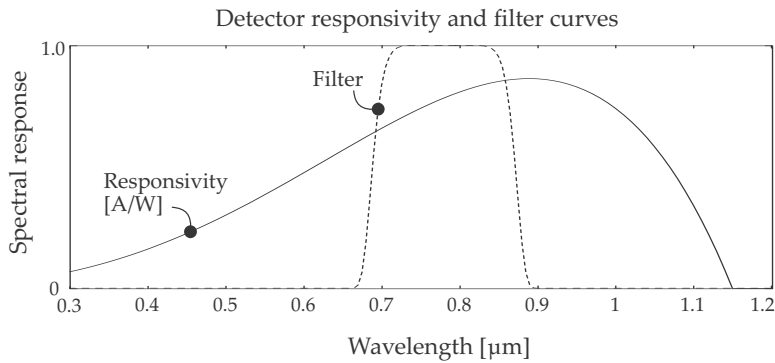
The sensor's InSb detector responsivity can be modeled by Equation (D.5) with ($\lambda_c = 6 \text{ }\mu\text{m}$, $k = 20$, $a = 3.5$ and $n = 4.3$).



- 6.7.1 Determine the instantaneous solid FOV of the detector through the lens, the projected solid angle of the lens from the detector's position, and the projected solid angle of the lens barrel from the detector's position. [3]
- 6.7.2 Calculate the current through the detector, assuming that the lens barrel temperature is maintained at 300 K. [2]
- 6.8 A square silicon detector is pointed toward the sun (normal vector is directed to the sun). The detector has an area of 1 cm^2 . You may assume unity atmospheric transmittance.

The sensor silicon detector responsivity can be modeled by Equation (D.5) with ($\lambda_c = 1.15 \text{ }\mu\text{m}$, $k = 8$, $a = 3.5$ and $n = 4.3$).

The sensor filter transmittance can be modeled by Equation (D.4) with ($\tau_s = 0$, $\tau_p = 1$, $\lambda_c = 0.8$, $\Delta\lambda = 0.21$ and $s = 100$).



- 6.8.1 Write a mathematical formulation of the problem; include flux transfer, detector response, etc. Describe all elements in the model and provide the relevant numerical values for all parameters. [5]
- 6.8.2 Implement the model in a computer program. Describe the structure of the model and provide all numeric values (spectral and scalar). Use a wavelength increment of $0.01 \mu\text{m}$. [5]
- 6.8.3 Calculate the current through the detector when it is viewing the sun, if no optical filter is present. [2]
- 6.8.4 Locate a lens with focal length of 100 mm and diameter of 35 mm in front of the detector, such that the detector is located in the focal plane of the lens (still no filter). Calculate the detector current for this condition, with the sensor facing the sun. [3]
- 6.8.5 Calculate the current through the detector plus lens when it is viewing the sun, with the optical filter present. [2]
- 6.8.6 The detector (without the lens and with no filter) is placed in an oven. The oven's inside dimensions are 0.5 m along all sides. Calculate the detector current for oven temperatures of 1300 K and 2800 K. Assume an oven wall emissivity of unity. [4]
- 6.8.7 The detector *with the lens* (but still no filter) is placed in the oven. Determine the detector currents for the two temperatures. Compare the detector current magnitudes with/without the lens. Describe the differences. [2]
- 6.8.8 Repeat the above four problems (facing the sun, with/without lens, facing the oven wall, with/without lens), but this time include the filter shown above in the same graph as the detector responsivity. [2]
- 6.8.9 Consider all of the results from the previous calculations. What conclusions can be drawn from these results? This is not a trick question, just think creatively and list all your observations. [4]
Evaluate your calculation method; was it suitable for these problems? [1]

Chapter 7

Radiometry Techniques

*When you can measure what you are speaking about,
and express it in numbers, you know something about it;
but when you cannot measure it, when you cannot express it in numbers,
your knowledge is of a meagre and unsatisfactory kind.*

William Thomson, Lord Kelvin

7.1 Performance Measures

Performance measures provide a quantified evaluation of system performance. These measures often indicate aggregate performance, covering several lower-level parameters in a single measure. Depending on the sensor application, one or more of these expressions provide an optimum sensor performance figure. For example, the sensitivity of an optical communication receiver is described in terms of the SNR ratio at the sensor, whereas the performance of a thermal camera is expressed in terms of the detectable temperature difference. The motivation for using performance measures is given in Section 1.2.8. Examples of the application of performance measures are given in Chapter 9.

7.1.1 Role of performance measures

Performance measures or ‘figures of merit’ are used to develop an understanding of a system’s performance. These are critical tools to be used during system design and for technical performance measurement during development phases. *The real value of performance measures lies not so much in the absolute value of the measure but in relative values when comparing two situations.* Performance measurement may be used as follows:

1. During the design phase, choices can be made by relative comparison of the predicted performance for different design options. The option with the highest predicted performance may then be further developed.

2. Performance measures can be used to optimize designs by trading off performance against design complexity.
3. During later development phases, performance measures can be used to monitor progress in technical performance against the planned progress.
4. Performance measures may be used to determine the sensitivity of a design to variations in component characteristics.
5. Performance measures may be used to determine the safety margins in a design.

7.1.2 General definitions

In the realist view, the *time* domain is a dimension with irreversible and sequential flow or movement from past, through present, into future. Time is described in a single real number relative to some fiducial epoch (starting event). When modeling time in mathematical terminology, time may appear to be reversible.

In the Cartesian view, the *space* domain is a three-dimensional extent in which objects have freedom of movement and direction. Space is commonly modeled in a coordinate system with the convention of using (x, y, z) to denote the three Cartesian dimensions. The origin of this coordinate system is uniquely defined in different problem spaces.

Foreground focus attaches positive importance or priority to an event or signal, in the context of its application. Foreground is the primary focus of an investigation.

Background focus attaches a neutral or negative priority to an event or signal, in the context of its application. Background interferes with the analysis of foreground. Background is everything that is not foreground.

Signal is a foreground continuous time or discrete function in time or space domains that conveys information.

Noise is a background continuous time or discrete function in time or space that interferes with the foreground function or event. Noise has a connection with unfolding time but can also be expressed in the space domain. There is a strong connotation that noise has mostly random behavior, but randomness is not a mathematical requirement (e.g., electromagnetic interference is regarded as a noise but is not random).

Clutter is a special type of noise, often with a connection to the geometry of the real world. Clutter is normally used in the context of image processing or radar processing, being caused by physical objects in the

problem spatial extent. Clutter is mostly structured noise, with little or no randomness. The (spatial/geometric) structure in clutter is what sets it apart from noise.

Root-mean-square (rms) is often used as a measure of noise. The rms of a function (normally a time signal) $f(t)$, with duration T , is defined as

$$f_{\text{rms}} = \lim_{T \rightarrow \infty} \sqrt{\frac{1}{T} \int_0^T [f(t)]^2 dt}. \quad (7.1)$$

The peak value is often used as an indication of instantaneous signal strength. A signal local peak or maximum is the value $f(x^*)$ over the range ϵ , where $f(x^*) \geq f(x)$ when $\|x - x^*\| < \epsilon$.

7.1.3 Commonly used performance measures

The measures listed below are widely used to quantify the performance of linear (or near-linear) systems. The list is not complete because many designs have their own unique performance measures. Designers must identify the relevant measures for their particular designs. Performance prediction and verification of nonlinear systems are not easily performed by simple measures such as these; nonlinear system performance is generally determined by using scenario-based simulation models or the real hardware item.

Signal-to-noise ratio (SNR): SNR is the ratio of the peak signal value to the rms value of the noise. Higher SNR values imply better system performance. SNR is unitless.

This measure depends on characteristics of the source, medium, and sensor. The noise includes sensor noise or photon noise in the source. It is normally easy to determine the SNR, both by measurement and calculation. The SNR is improved by optimizing the bandwidth, increasing source intensity, or decreasing sensor noise.

Signal-to-clutter ratio (SCR): SCR is the ratio of the peak signal value to the peak clutter value. Higher SCRs imply better system performance. SCR is unitless.

This measure depends on the characteristics of the surroundings of the foreground target object. The clutter could be terrain objects reflecting ambient light or infrared radiation from hot terrain objects such as rocks. The medium and sensor have only limited effect on the SCR. The SCR is often measured because it is difficult to calculate in analytical form. The SCR is improved by increasing source intensity or optimizing the sensor FOV.

Noise equivalent power (NEP): NEP is the optical signal power required to create an electronic signal such that the SNR is one. Lower values of NEP imply a better performance. NEP has units of [W].

NEP requires the quantification and exact definition of noise and signal. NEP considers the total wideband noise (not power spectral density), and hence the bandwidth must be specified. Likewise, the properties of the source and medium determine the signal and must be specified.

NEP can be defined for a complete sensor or just a detector on its own. NEP is mostly calculated because it is difficult to measure directly. The NEP is improved by decreasing the various noise sources. See also Section 6.7.

Detectivity (D): Detectivity is the reciprocal of NEP. Detectivity is the SNR for a 1-watt input signal (normally a very big number!). Detectivity has units of [W⁻¹].

Specific detectivity, ‘Dee-star’ (D*): D* is the noise in a detector scaled with noise bandwidth and detector area. The D* describes the detector performance in fundamental or absolute terms. It is commonly quoted in product data sheets. D* is normally specified in units of [cm·√Hz/W].

Wideband, as well as spectral, D* and NEP are related by

$$NEP = \frac{\sqrt{\Delta f A_d}}{D^*}, \quad (7.2)$$

where Δf is the noise equivalent bandwidth, and A_d is the detector area.

Noise equivalent irradiance (NEE): NEE is the optical irradiance required to create an electronic signal such that the SNR is one. Lower values of NEE imply a better performance. NEE is an expression of sensor noise but is affected by the source properties and the optical medium (atmosphere). NEE has units of [W/m²].

The NEE is the NEP (see above) scaled by the effective sensor aperture area and filter/optics transmittance. See also Section 6.7.

Noise equivalent radiance (NEL): NEL is the radiance (from an extended source) required to create an electronic signal such that the SNR is one. Lower values of NEL imply a better performance. NEL is an expression of sensor noise but is affected by the source properties and the optical medium (atmosphere). NEL has units of [W/(m²·sr)]. See also Section 6.7.

Noise equivalent exitance (NEM): NEM is the exitance (from an extended source) required to create an electronic signal such that the SNR is one. Lower values of NEM imply a better performance. NEM is an expression of sensor noise but is affected by the source properties and the optical medium (atmosphere). NEM has units of $[W/m^2]$. See also Section 6.7.

Noise equivalent temperature difference (NETD or $NE\Delta T$): NETD is the temperature difference (from an extended source) required to create a SNR of one. A lower NETD implies better performance. NETD is an expression of sensor noise but is affected by the source properties and the optical medium (atmosphere). NETD is derived from instrument measurements; the human observer is not part of this measurement. It is usually measured at a background temperature of 300 K. NETD has units of [K]. See also Sections 6.7 and 9.5.4.

Minimum resolvable temperature (MRT): MRT is the smallest temperature difference between two (extended source) blackbodies (reference blackbody at 300 K), arranged in a standard bar-chart test pattern with 4 line pairs, that can be observed by humans. MRT has units of [K]. MRT is a performance figure that includes the human observer as part of the measurement.

For design purposes the human observer is sometimes modeled with a set of equations and parameters.¹⁻⁴ There is not always agreement on these models, and care must be taken when comparing MRT values quoted by different sources.

Minimum detectable temperature (MDT): MDT is the smallest temperature difference between two blackbodies, arranged as a rectangular plate against a background at 300 K, that can be observed by humans. MDT has units of [K]. The same comments as for MRT apply.

Noise equivalent reflectance (NER or $NE\Delta\rho$): NER is the change in reflectance required to create an electronic signal such that the SNR is one. Lower values of NER imply a better performance. NER is unitless. The spectral band and the source spectral radiance used to illuminate the surface must be specified.

Probability of detection (P_d): For a signal corrupted by noise, P_d is the probability that the signal will exceed a (fixed or variable) threshold. A higher probability of detection implies better performance. P_d is unitless, with $0 \leq P_d \leq 1$. Probability of detection is a three-fold function of signal magnitude, noise/clutter magnitude, and threshold.

It is relatively difficult to measure P_d as it approaches zero or unity because the statistical events become relatively rare. P_d is improved by lowering the threshold relative to the noise and clutter magnitudes, but a lower threshold results in increased false alarm rate. Simplified noise calculations can be done, but accurate calculation of P_d is quite difficult.^{2,5} See also Section 7.8.

Probability of false detection (P_n): For a signal corrupted by noise, P_n is the probability that the noise will exceed a (fixed or variable) threshold when no signal is present. A lower probability of false detection implies better performance. P_n is unitless. Note that $0 \leq P_n \leq 1$ and is usually small in a well-designed system.

The probability of false detection is a function of the noise level and threshold setting. P_n is improved by increasing the threshold relative to the noise and clutter magnitudes. See also Section 7.8.

False alarm rate (FAR): For a signal corrupted by noise, the false alarm rate is the rate by which the threshold is exceeded in the absence of a signal. FAR is unitless. The false alarm rate is related to P_n by

$$FAR = \frac{NP_n}{t_d}, \quad (7.3)$$

where N is the number of detectors, and t_d the time on target.

Point spread function (PSF): PSF is the impulse response of an optical system: the flux distribution in the image plane of a point source object.⁶ The PSF results from lens aberrations and diffraction. It is therefore a property of the optics, but some investigations could include medium effects such as turbulence. Smaller PSFs imply better performance. In imaging systems the PSF is an indication of the sharpness of the image, and in nonimaging systems the PSF is an indication of the amount of power that will fall onto a detector. The PSF can be accurately calculated by modern optical-design computer programs. See also Section 6.3.7.

Optical transfer function (OTF): The OTF is the Fourier transform of the optical PSF. It is a complex, two-dimensional function. Bigger volumes under the OTF generally imply better performance.

Modulation transfer function (MTF): The MTF is the absolute value of the Fourier transform of the optical PSF. It is a real, two-dimensional function. Bigger volumes under the MTF generally imply better performance. The MTF can be considered as the 'spatial frequency response' of the lens, i.e., the lens' ability to convey the high-frequency

spatial information in an image. MTF is very commonly used by optical and electro-optical system designers. Care should be exercised when interpreting only the tangential and sagittal sections through the MTF of a lens because incorrect conclusions can easily be drawn.

7.2 Normalization

Normalization^{7,8} is the process whereby a function (e.g., spectrally varying parameter) is reduced to a set of simple numbers, such as 'effective bandwidth' or 'average responsivity.' *Normalization removes information* from the initial data set; it should only be used if the user understands the process whereby the normalization was achieved.

7.2.1 Solid angle spatial normalization

Spatial normalization has already been encountered when considering spatial solid angles and view factors in Sections 2.5 and 2.8. In effect, the arbitrarily complex spatial distribution of a surface is reduced to a simple number.

7.2.2 Effective value normalization

The effective value of a variable is given by

$$\mathcal{F}_{\text{eff}} = \frac{\int_0^\infty \mathcal{F} \mathcal{G} d\lambda}{\int_0^\infty \mathcal{G} d\lambda}, \quad (7.4)$$

where \mathcal{F} is the variable in question, and \mathcal{G} is a weighting function. Note that the effective value of \mathcal{F} depends on the shapes of both \mathcal{F} and \mathcal{G} ; the effective value of \mathcal{F} thus calculated therefore applies only to the specific weighting function \mathcal{G} .

Effective value normalization is discussed in Section 4.4, in the calculation of effective transmittance. The sensitivity of effective transmittance to CO₂ spectral transmittance was demonstrated.

Effective value normalization also appears in the calculation of detector wideband responsivity. The detector with area A_1 is illuminated with unfiltered thermal body radiation with area A_0 at a temperature T_c , usually 500 K or 1000 K. The wideband responsivity is now defined as the ratio of the signal current [Equation (6.16), with $Z_t = 1$ and $k = 1$] to the

total irradiance onto the detector:

$$\begin{aligned}\mathcal{R}_I &= \frac{i_s}{\Phi_s} = \frac{\frac{A_0 A_1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda}(T_c) \tau_{a\lambda} \mathcal{R}_\lambda d\lambda}{\frac{A_0 A_1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda}(T_c) \tau_{a\lambda} d\lambda} \\ &= \frac{\int_0^\infty \epsilon_{0\lambda} L_{0\lambda}(T_c) \tau_{a\lambda} \mathcal{R}_\lambda d\lambda}{\int_0^\infty \epsilon_{0\lambda} L_{0\lambda}(T_c) \tau_{a\lambda} d\lambda}.\end{aligned}\quad (7.5)$$

The effective responsivity depends on the spectral shapes of both the detector and the apparent source spectral radiance (as observed through the atmosphere). The responsivity value thus calculated applies only to the source at the particular temperature used in the calculation.

To illustrate the sensitivity of Equation (7.5) to the source temperature, a simple calculation was performed. A spectral detector responsivity \mathcal{R}_λ was calculated using Equation (D.5) with $k = 1$, $a = 1$, $n = 15$, and $\lambda_c = 6.1$. Effective responsivity values were calculated for thermal radiator sources at 900 K and 1000 K. The effective responsivity at 900 K is 0.56, whereas the effective responsivity at 1000 K is 0.54. The two effective responsivity values differ by 4% for a temperature difference of only 10%. Suppose now that this detector is used to detect a source of 500 K; what good is any of these responsivity values?

7.2.3 Peak normalization

Peak normalization calculates

$$\Delta x = \frac{1}{\max(f)} \int_0^\infty f(x) dx, \quad (7.6)$$

where $f(x)$ is a function, and $\max(f)$ is the maximum value of the function. One example of peak normalization is the noise equivalent bandwidth of an electronic filter (see Section 5.3.13). Figure 7.1 shows the filter gain and the ratio between the noise bandwidth and the -3 dB electronic bandwidth for a class of filters known as Butterworth filters.

The noise equivalent bandwidth of an ideal *switched integrator* is given by

$$\Delta f = \int_0^\infty \left(\frac{\sin \pi T f}{\pi T f} \right)^2 df \quad (7.7)$$

$$= \frac{1}{2T}, \quad (7.8)$$

where T is the integration time of the integrator. The -3 dB bandwidth of the integrator is $f_{-3\text{ dB}} = 1/(2.273 T)$. The ratio between the noise equivalent bandwidth and the -3 dB bandwidth is 1.156.

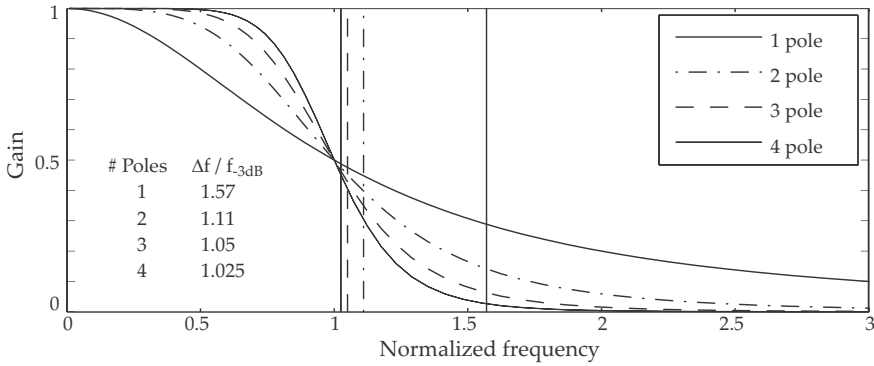


Figure 7.1 Butterworth filter noise equivalent bandwidth for different filter orders.

7.2.4 Weighted mapping

The weighted value of a variable is given by

$$\mathcal{F}_{\mathcal{G}} = \int_0^{\infty} \mathcal{F} \mathcal{G} d\lambda, \quad (7.9)$$

where \mathcal{F} is the variable in question, and \mathcal{G} is a weighting function. Note that the weighted value of \mathcal{F} depends on the shapes of both \mathcal{F} and \mathcal{G} . The result is a scalar expression of \mathcal{F} after it is mapped into a new space by \mathcal{G} . Essentially, the intent with weighted mapping is to express the relevancy of one function in terms of the space defined by another function. If there is little overlap between the vectors, the result will be small.

Section 2.10.4 introduces weighted mapping of spectral radiance to color coordinates. This particular weighting accounts for the spectral response of a sensor — in this case, the human eye. It is evident that weighted mapping provides a technique to weigh or ‘measure’ a spectral variable for relevancy for a particular sensor’s spectral response.

Equation (6.19) defined the ‘inband’ irradiance as the spectral irradiance weighted by the sensor’s spectral response

$$E_S = \frac{k d A_0 \cos \theta_0}{R_{01}^2} \int_{\lambda_1}^{\lambda_2} (\epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda}) (\mathcal{S}_{\lambda}) d\lambda,$$

where $\epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda}$ is the source radiance after transmission through the atmosphere, and \mathcal{S} is the weighting function. In this case, the apparent source radiance is re-mapped to the sensor’s system response \mathcal{S} to represent the radiance as observable by the sensor.

7.3 Spectral Mismatch

Spectral calculations are performed with variants of Equations (2.33) and (6.16), repeated here for convenience:

$$v_S = \frac{k \hat{R} Z_t dA_0 \cos \theta_0 A_1}{R_{01}^2} \int_{\lambda_1}^{\lambda_2} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} S_\lambda d\lambda, \quad (7.10)$$

where the spectral system response is given by S_λ . In practice, the system response is defined by filters, optical elements, and the detector. The manufactured devices have statistical spread in transmittance curves. Some components may extend a little toward the shorter wavelengths, whereas other components extend a little toward the longer wavelengths — even though all components are within specification. Consider now an acceptance test setup that is used to evaluate a batch of sensors. For a given source signal, the sensor must provide a certain minimum signal level to pass the test.

If the sensors are evaluated with a source heavily weighted toward, say, the longer wavelengths in the sensor spectral band, the effect of filter variations in the longer wavelengths are accentuated, whereas the effect of variations in the shorter wavelengths are less accentuated. If the sensor is to be used operationally with a different source spectrum, the acceptance test could be considered invalid. The test would be invalid because the test source spectrum over the full sensor spectral band does not represent the actual operational target spectrum. This scenario could occur for very sensitive sensors, requiring very low signal levels. The requirement for such evaluations is that a low signal radiance is required but also at the correct spectral shape over the full sensor band.^{9,10}

For example, assume a sensor that must be evaluated in the 3–5- μm spectral band at an irradiance level of $0.1 \mu\text{W}/\text{m}^2$ with a source spectrum typical of a 700-K blackbody. If the test equipment employs industry-standard sources and collimators, the achievable irradiance levels are some orders of magnitude too high. One way out of this situation is to lower the source temperature to such a value (say to 400 K) that the required irradiance is achieved. At this low temperature, the source spectrum is very different from the required 700-K source spectrum; it would be strongly weighted toward the longer wavelengths. Evaluation tests on this test setup will not be representative of the sensor's response to a 700-K source.

7.4 Spectral Convolution

Section 6.6 describes how to calculate the signal that a given sensor with response \mathcal{S} would receive from a source L , through some medium with transmittance $\tau_{0\lambda}$. Suppose that this sensor has a spectral filter τ_f with a narrow (but nonzero) spectral width. This filter is very narrow compared with its central wavelength, say $\Delta\lambda = 0.01\lambda_c$. Such radiometers are used to determine the spectral radiance of sources or to measure the spectral transmittance of the atmosphere. The apparent irradiance measured by such a system can be written [from Equation (6.13)]

$$E_{\lambda_c} = k \int_0^\infty \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \tau_{f\lambda} \mathcal{S}_\lambda d\lambda, \quad (7.11)$$

where k accounts for the geometrical factors such as source area, orientation, and distance. This equation can be written as

$$E_\lambda = k \int_{\lambda_c - \frac{\Delta\lambda}{2}}^{\lambda_c + \frac{\Delta\lambda}{2}} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \tau_{f\lambda} \mathcal{S}_\lambda d\lambda. \quad (7.12)$$

By change of variable $\lambda = \lambda_c - x$,

$$E_{\lambda_c} = k \int_{-\frac{\Delta\lambda}{2}}^{+\frac{\Delta\lambda}{2}} \epsilon_{0x} L_{0x} \tau_{ax} \tau_{f(\lambda_c-x)} \mathcal{S}_x dx. \quad (7.13)$$

These equations show very clearly that the irradiance measured with the filter centered around wavelength λ_c includes source energy from $\lambda_c - \frac{\Delta\lambda}{2}$ to $\lambda_c + \frac{\Delta\lambda}{2}$. Apart from the spectral selection, the filter has an additional effect by smoothing the spectrum being observed because the filter has a nonzero spectral width.

Equation (7.13) is called a convolution integral because it describes the convolution between the product $(\epsilon_{0\lambda_c} L_{0\lambda_c} \tau_{a\lambda_c} \mathcal{S}_{\lambda_c})$ and τ_f . In linear systems terminology, the observed spectral source radiance is being convolved with the filter spectral transmittance. To investigate the effects of this convolution consider the two cases: (1) the observed source has little variation over the filter passband, and (2) the observed source varies significantly over the filter passband.

If the product $(\epsilon_{0\lambda_c} L_{0\lambda_c} \tau_{a\lambda_c} \mathcal{S}_{\lambda_c})$ is more or less constant over the filter passband $\Delta\lambda$, Equation (7.13) can be written to show that the convolution has little effect other than some insignificant amount of smoothing:

$$E_{\lambda_c} = k (\epsilon_{0\lambda_c} L_{0\lambda_c} \tau_{a\lambda_c} \mathcal{S}_{\lambda_c}) \int_{-\frac{\Delta\lambda}{2}}^{+\frac{\Delta\lambda}{2}} \tau_{f(\lambda_c-x)} dx \quad (7.14)$$

$$\approx k \epsilon_{0\lambda_c} L_{0\lambda_c} \tau_{a\lambda_c} \mathcal{S}_{\lambda_c} \tau_f \Delta\lambda. \quad (7.15)$$

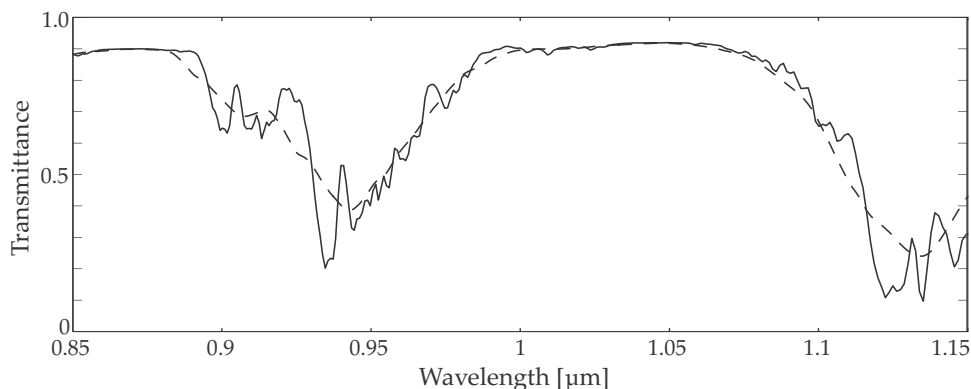


Figure 7.2 Atmospheric transmittance convolved with 10 cm^{-1} and 300 cm^{-1} square windows.

If the product $(\epsilon_{\lambda} L_{\lambda} \tau_{a\lambda} S_{\lambda})$ varies significantly over the filter passband $\Delta\lambda$, Equation (7.13) cannot be simplified. In this case, the convolution attenuates and smears out the finer detail in the spectral information. If the actual spectral line is very narrow the measured line will approximate the filter resolution and will be totally erroneous unless the filter spectral smear effect is compensated by deconvolution.

The effect is best illustrated in Figure 7.2. MODTRANTM was used to calculate the spectral transmittance in the $0.85\text{--}1.1\text{-}\mu\text{m}$ spectral range. The calculation was performed for a 1-km path length at sea level in a Tropical atmosphere at 27°C , 75 \%RH , 1015 mB , 23-km visibility. The transmittance data so obtained was convolved with a square filter response. The spectral filter widths are 10 cm^{-1} and 300 cm^{-1} .

As an example of one effect of the finite filter width, consider the case where a spectral source is observed with a spectral radiometer. The estimated source spectrum (a) and measured spectral curve (b) are shown in Figure 7.3. Suppose MODTRANTM was used to calculate the spectral transmittance (c) of the atmosphere. In order to compensate for the atmospheric attenuation, the measured irradiance is divided by the transmittance curve, resulting in the calculated spectrum (d). The sequence on the left in Figure 7.3 shows what happens if the atmospheric transmittance was calculated at a much-higher resolution than the spectral measurement filter width. In the sequence on the right in Figure 7.3, the atmospheric transmittance (c) was obtained by convolving the high-resolution MODTRANTM data with the filter response before dividing the measured source irradiance with the atmospheric transmittance.

In order to prevent serious errors in narrow-band spectral calculations, all of the spectral variables must first be convolved to the same spectral

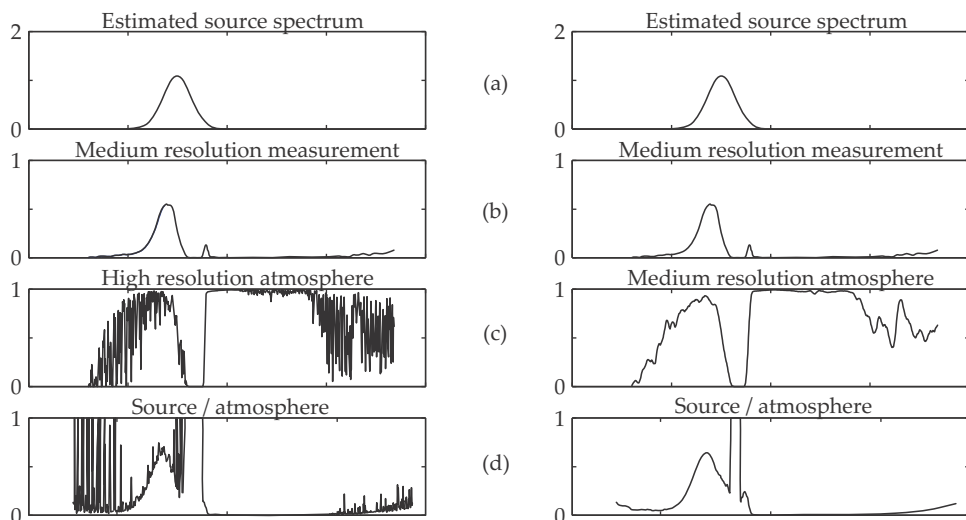


Figure 7.3 Atmospheric correction of radiometer measurements in various stages of processing.

resolution before spectral multiplications or divisions are performed. If the spectral variables are not all convolved to the same basic resolution, one can obtain erroneous spectra or emissivity values, or transmittance values exceeding unity.

7.5 The Range Equation

It is frequently necessary to determine the operational detection distance of a source and sensor combination. The problem is usually stated as follows: “What operating detection range can be achieved with a given source intensity, atmospheric attenuation, and sensor sensitivity?” The objective is to solve for R in

$$E = \frac{I\tau_a(R)}{R^2}. \quad (7.16)$$

Rewrite this equation as

$$\frac{I}{E} = \frac{R^2}{\tau_a(R)}, \quad (7.17)$$

where the left side is a constant given by the source intensity I and threshold irradiance E , whereas the right side describes the range-related terms. In most cases the solution is not simple and requires an iterative numerical solution.^{11,12} A numerical solution is shown in Section D.5.6.

The range equation can also be solved graphically by plotting the ratio I/E for various atmospheric conditions, as function of range. This is

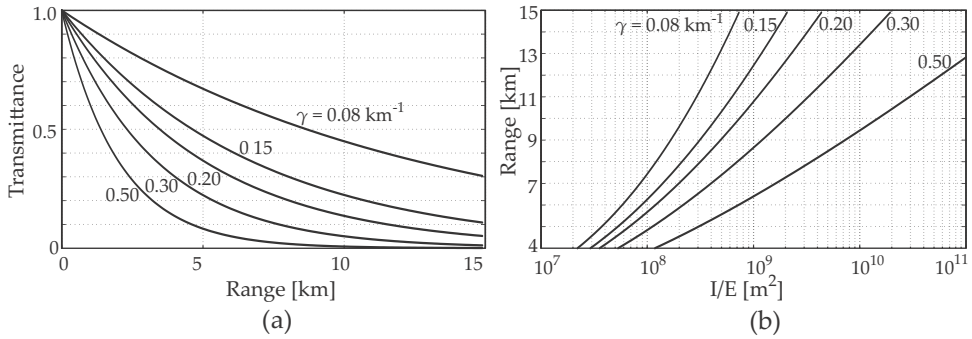


Figure 7.4 Determining range from the range equation.

shown on the right side in Figure 7.4. The vertical lines are for various I/E values. The intercepts of the vertical I/E lines with the $R^2/\tau_a(R)$ curves indicate the detection ranges.

7.6 Pixel Irradiance in an Image

An imaging system, here referred to as the *observer*, forms a two-dimensional representation of the scene called an *image*. The image consists of a number of smaller picture elements called *pixels*. The pixel magnitude is proportional to the radiation received within the solid angle subtended by each pixel. The atmosphere attenuates the scene radiance and adds path radiance. The pixel signal magnitude (irradiance or detector voltage) is a function of object size, range, and the relative background contribution. In this section an analytical description of the relationship between pixel signal magnitude and range to a given object, against a given background, is derived.

If the solid angle subtended by the object is smaller than the pixel solid angle, the object is said to be *unresolved*. If the solid angle subtended by the object is significantly bigger than the pixel solid angle, the object is said to be *resolved* against an extended target. For resolved objects, the atmosphere degrades the contrast in the scene, whereas for unresolved objects the pixel signal magnitude is reduced with increasing range, as well as degraded by the atmosphere. This analysis ignores sky- and ground-induced path radiance. As discussed in Section 4.2.2, this means that the current derivation is only valid for horizontal paths.

Consider first a target object at some range R_{T1} from the observer, with projected area $A'_T = A_T \cos \theta_T$ (limited to the pixel FOV), and radiance L_T . The atmospheric transmittance between the target object and observer is τ_T . Consider second the background, around or behind the target object,

with projected area $A'_B = A_B \cos \theta_B$ (limited to the pixel FOV), at range R_{B1} , the background has a radiance L_B , with atmospheric transmittance to the observer of τ_B . The background and the object may not be at the same distance. The atmospheric path radiance in the direction of the object is L_P . The observer has an instrument with optical aperture A_1 , and it measures irradiance $E_1 = d\Phi/dA_1$. The pixel irradiance consists of four components: the object irradiance, path irradiance in front of the object, background irradiance, and path irradiance in front of the background:

$$dE_1 = \frac{L_T A'_T \tau_T}{R_{T1}^2} + \frac{L_B A'_B \tau_B}{R_{B1}^2} + \frac{L_{Pt} A'_T}{R_{T1}^2} + \frac{L_{Pb} A'_B}{R_{B1}^2} \quad (7.18)$$

$$= L_T \Omega_T \tau_T + L_B \Omega_B \tau_B + L_{Pt} \Omega_T + L_{Pb} \Omega_B, \quad (7.19)$$

where $\Omega_T = A'_T/R_{T1}^2$ is the object projected solid angle (limited to the pixel FOV), $\Omega_B = A'_B/R_{B1}^2$ is the background projected solid angle (limited to the pixel FOV), L_{Pb} is the path radiance for the fraction of the pixel filled by the background, and L_{Pt} is the path radiance for the fraction of the pixel filled by the target object. In the analysis shown here, L_A is the atmospheric radiance (as applicable to path radiance, see Section 4.2.2). The projected pixel footprint area $A_P \cos \theta_P = \Omega_P R^2$ is defined as the projected area that the pixel subtends at a given range R . This scenario is further investigated in Sections 9.2 and D.5.3.

The pixel magnitude as a function of object-to-observer range is shown in Figure 7.5. Depending on the object size and range, Equation (7.18) can be recast into several different forms. At ranges corresponding to region I the object is resolved (object larger than the pixel footprint), whereas at ranges corresponding to regions II to IV the object is unresolved (object smaller than pixel footprint).

Region I

The target completely fills the pixel FOV, and no background is visible. For a resolved object $A_P \cos \theta_P \leq A_T \cos \theta_T \Rightarrow A_B = 0$, then $A_P \cos \theta_P / R^2 = \Omega_P$ and

$$E_1 = \frac{L_T A_P \cos \theta_P \tau_T}{R_{T1}^2} + L_{Pt} \Omega_P \quad (7.20)$$

$$= (L_T \tau_T + L_{Pt}) \Omega_P$$

$$\approx [L_T e^{-\gamma R_{T1}} + L_A (1 - e^{-\gamma R_{T1}})] \Omega_P. \quad (7.21)$$

For a resolved object the pixel magnitude is given by the object radiance multiplied by the atmospheric transmittance term. Because the atmospheric transmittance is a function of range, the object magnitude will decrease with $e^{-\gamma R}$, whereas the path radiance increases with $(1 - e^{-\gamma R})$ [see Equation (4.10)].

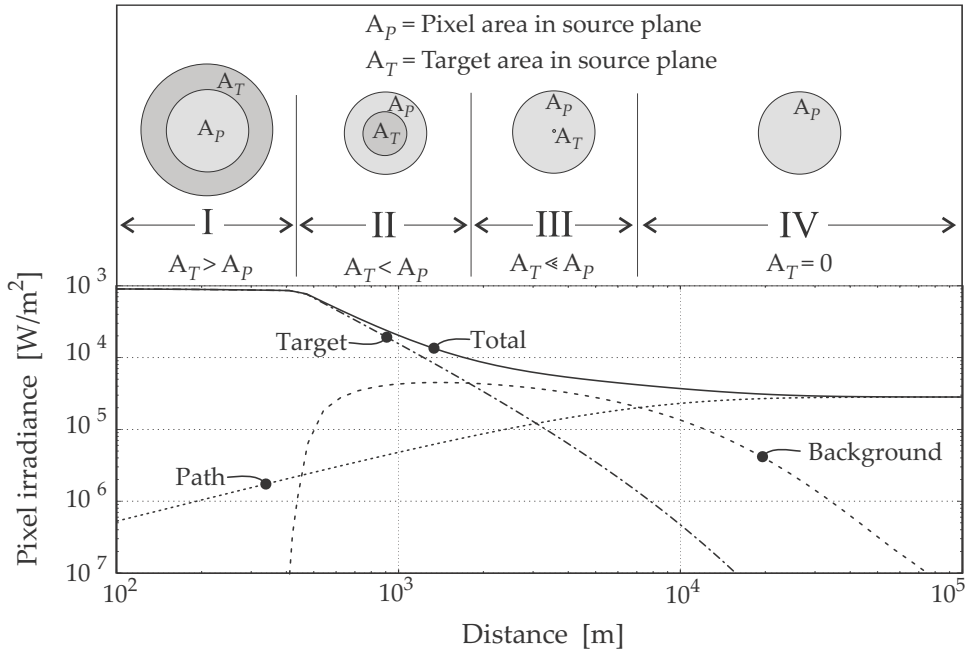


Figure 7.5 Pixel magnitude as a function of object-to-observer distance, as affected by the atmosphere.

In this region, the object signal is only affected by atmospheric effects because the object is large and observed at (relatively) close range. When moving away from the object (i.e., for increasing R), the FOV is constantly filled by new target area; hence the $1/R^2$ free space loss is offset by observing more of the target area. As the object range increases, the atmosphere attenuates the object radiance.

Region II

The target partially fills the pixel FOV. Then, $A_P \cos \theta_P > A_T \cos \theta_T \Rightarrow A_B \neq 0$. Now, $\Omega_T = A_T \cos \theta_T / R_{T1}^2$, so that with increasing range, Ω_T decreases, and the background starts filling around the target in the pixel FOV. Define the solid angle subtended by the background as $\Omega_B = \Omega_P - \Omega_T = \Omega_P - A'_T / R_{T1}^2$, then

$$E_1 = L_T \Omega_T \tau_T + L_B \Omega_B \tau_B + L_{Pt} \Omega_T + L_{Pb} \Omega_B \quad (7.22)$$

$$= [L_T e^{-\gamma R_{T1}} + L_A (1 - e^{-\gamma R_{T1}})] A'_T / R_{T1}^2 + [L_B e^{-\gamma R_{B1}} + L_A (1 - e^{-\gamma R_{B1}})] (\Omega_P - A'_T / R_{T1}^2). \quad (7.23)$$

In region II, the variation of the four flux components is a complex function of R_{T1} , as is evident in Equation (7.23). In summary, the object and the path irradiance in front of the object decrease with

the reciprocal of range, whereas the background irradiance increases concomitantly.

Region III

The target fills only a very small portion of the pixel FOV, and the background is the dominant source. For $L_B\Omega_B \gg L_T\Omega_T$, or Ω_T very small:

$$\begin{aligned} E_1 &= L_B\Omega_P\tau_B + L_{pb}\Omega_P \\ &= [L_B e^{-\gamma R_{B1}} + L_A(1 - e^{-\gamma R_{B1}})]\Omega_P. \end{aligned} \quad (7.24)$$

At long range, the object flux decrease attributable to $1/R^2$ loss is so severe that the object signal is less than the background signal, so that the pixel contains mainly background and path radiance flux. In region III the irradiance equation has the same form as in region I, except that the target object is now replaced with the background.

Region IV

At longer range, even the background radiance is severely attenuated and the only remaining flux is due to path radiance. Hence, for $L_B\tau_B \ll L_P$, the image pixel irradiance is given by

$$E_1 = L_P\Omega_P. \quad (7.25)$$

At very long ranges the path radiance dominates all other sources, and the object and clutter are lost in the 'fog-like' haze caused by atmospheric radiance.

7.7 Difference Contrast

The radiance of an object in its surroundings consists of a global constant radiance level and a local radiance variation. Some electro-optical systems are only sensitive to the local variation in the radiance — called the difference contrast — between the object and its surroundings. The constant radiance level is removed by filtering or some other means.

Consider a scene at uniform background temperature T_B containing one object at a temperature T_O . The spectral irradiance at the sensor due to a pixel filled only by the background is

$$E_{B\lambda} = L_{B\lambda}\Omega_P\tau_{B\lambda} + L_{pb\lambda}\Omega_P, \quad (7.26)$$

where $L_{B\lambda}$ is the background radiance, Ω_P is the sensor pixel FOV, $\tau_{B\lambda}$ is the atmospheric transmittance between the sensor and the background, and $L_{pb\lambda}$ is the path radiance.

If a pixel is only partially filled by the target object, the spectral irradiance at the sensor is

$$E_{T\lambda} = (L_{T\lambda}\tau_{T\lambda} + L_{P_{t\lambda}})\Omega_T + (L_{B\lambda}\tau_{B\lambda} + L_{P_{b\lambda}})(\Omega_P - \Omega_T), \quad (7.27)$$

where $L_{T\lambda}$ is the object radiance, Ω_T is the solid angle subtended by the object, $L_{P_{t\lambda}}$ is the path radiance in front of the object, and $\tau_{T\lambda}$ is the atmospheric transmittance between the object and the background.

The local radiance variation, or radiometric contrast, is now the difference between the pixel filled with the object and the pixel filled by background only:

$$\begin{aligned} \Delta E_\lambda &= (L_{T\lambda}\tau_{T\lambda} + L_{P_{t\lambda}})\Omega_T + (L_{B\lambda}\tau_{B\lambda} + L_{P_{b\lambda}})(\Omega_P - \Omega_T) \\ &\quad - L_{B\lambda}\Omega_P\tau_{B\lambda} - L_{P_{b\lambda}}\Omega_P \\ &= L_{T\lambda}\tau_{T\lambda}\Omega_T + L_{P_{t\lambda}}\Omega_T - L_{B\lambda}\tau_{B\lambda}\Omega_T - L_{P_{b\lambda}}\Omega_T \\ &= (L_{T\lambda}\tau_{T\lambda} - L_{B\lambda}\tau_{B\lambda})\Omega_T + (L_{P_{t\lambda}} - L_{P_{b\lambda}})\Omega_T. \end{aligned} \quad (7.28)$$

If the target object and background are at the same range, $\tau_{T\lambda} = \tau_{B\lambda}$ and $L_{P_{t\lambda}} = L_{P_{b\lambda}}$:

$$\Delta E_\lambda = (L_{T\lambda} - L_{B\lambda})\tau_{T\lambda}\Omega_T, \quad (7.29)$$

hence the signal is proportional to the difference in the radiance values between the object and the background.

7.8 Pulse Detection and False Alarm Rate

The detection of pulse signals corrupted by noise forms the basis of many electro-optical systems. The calculation of the probability of detection and false alarm rate requires information about the peak signal, rms noise, and threshold setting. The general solutions¹³⁻¹⁵ for arbitrary pulse shapes, noise spectra, and nonlinear detection processes are very complex and are not considered here. If the signal is a square pulse, filtered by a matched filter, and the input noise is white with a Gaussian distribution, as is the case for most natural noises, the detection performance can be readily^{13,16} calculated. This is a special case, but, is useful to obtain at least an order of magnitude indication.

The detection of a square pulse of width t_p , immersed in white noise after passing through a matched filter with bandwidth $\Delta f = 1/(2t_p)$, is shown graphically in Figure 7.6. Detection is the event where the signal corrupted by noise exceeds the detection threshold. A false alarm is the event where the noise (with no signal present) exceeds the detection threshold. The average false alarm rate is given by

$$FAR = \frac{1}{2t_p\sqrt{3}} \exp^{-t_i^2/(2t_p^2)}, \quad (7.30)$$

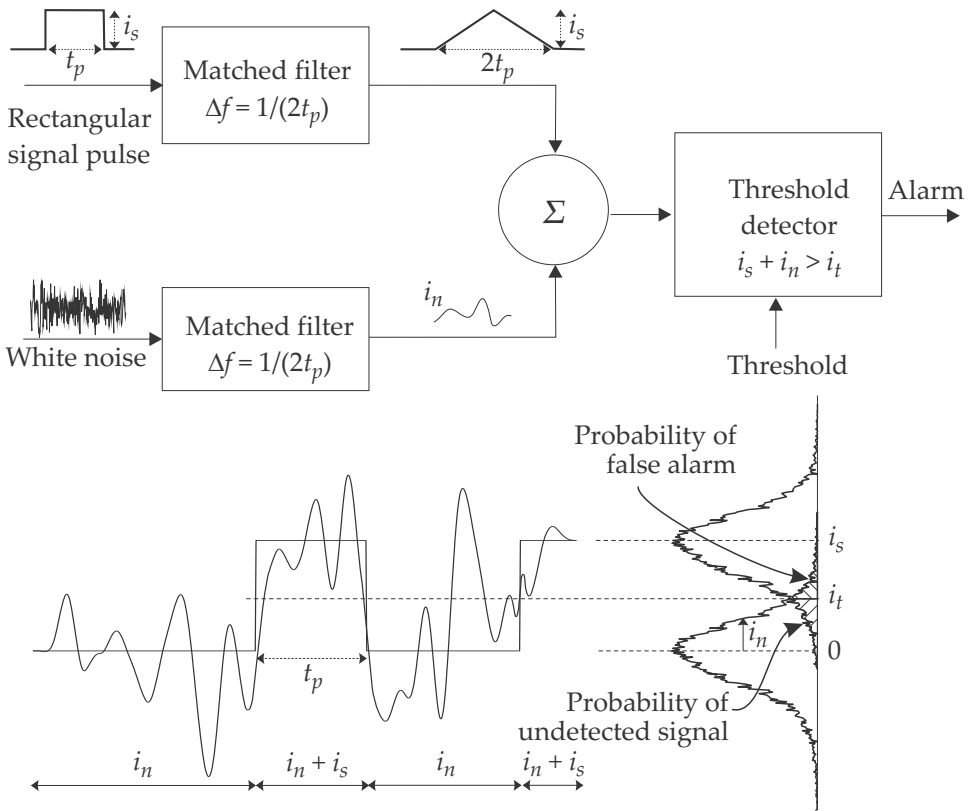


Figure 7.6 Diagram of the noise, signal, and threshold.

where i_t is the threshold value, and i_n is the rms noise value at the input to the threshold detector.

When there is a signal present, the probability of detection (signal plus noise exceeds the threshold) is given by

$$P_d \approx \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{i_s - i_t}{\sqrt{2}i_n} \right) \right], \quad (7.31)$$

where erf is the error function:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \quad (7.32)$$

For a given combination of threshold value i_t , noise value i_n , and signal value i_s , the average false alarm rate and probability of detection can be determined. Alternatively, for a given required false alarm rate or probability of detection, the threshold-to-noise ratio and SNR can be determined. These equations can be solved numerically or graphically.¹⁶

As an example of the application of these formulae, consider the following problem:¹⁶ The SNR and threshold-to-noise ratio (TNR) for a laser rangefinder must be found. The laser pulses are 100 ns wide. A range gate of 67 μ s is used. One in 1000 pulses may be lost, and a false alarm detection of 1 in 1000 pulses is required. The solution is as follows:

1. The false alarm performance must be 1 per 1000 pulses, each pulse arriving in a 67- μ s window. The false alarm rate must therefore be less than $1/(1000 \times 67 \times 10^{-6}) = 15$ pulses per second.
2. $t_p FAR = 0.1 \times 10^{-6} \times 15 = 1.5 \times 10^{-6}$.
3. The TNR is calculated as

$$\begin{aligned} \frac{i_t}{i_n} &= \sqrt{-2 \log_e (2t_p \sqrt{3} FAR)} \\ &= \sqrt{-2 \log_e (2\sqrt{3} 1.5 \times 10^{-6})} \\ &= 4.93. \end{aligned} \quad (7.33)$$

4. The SNR is given by

$$\frac{i_s}{i_n} = \sqrt{2} \operatorname{erf}^{-1} (2P_d - 1) + \frac{i_t}{i_n}, \quad (7.34)$$

requiring a SNR of 8.023.

The equations can be solved by the code given in Section D.5.7.

When considering the probability of detection in an image, each pixel is regarded as a separate channel, with a pulse width equal to its integration period. To convert from the false alarm rate per image to the false alarm rate per pixel, divide the required system false alarm rate by the number of pixels:

$$FAR_p = \frac{FAR_s}{N_p}. \quad (7.35)$$

The key design parameters are SNR (determines the probability of detection) and the signal-to-threshold ratio (determines the false alarm rate). In order to achieve high probability of detection, the designer strives toward a high SNR.

Consider an infrared point target detection system locating hot spots in a 4×10^5 -pixel image. The system is designed to operate at a SNR ratio of 12. Using Equation (7.31) it is found that for $\log(t_p FAR) = -22$,

the probability of detection is 98%. This operating point, for a sensor integration time of 1 ms, corresponds roughly with a FAR of 1×10^{-19} per pixel, or 4×10^{-14} per frame. This corresponds to roughly 4×10^{-9} false alarms per hour. Again, suppose the system design is done for a SNR ratio of 12. Using Equation (7.31) it is found that for $\log(t_p \text{FAR}) = -15$, the probability of detection is 99.99%. This operating point, for a sensor integration time of 1 ms, corresponds roughly with a FAR of 1×10^{-12} per pixel, or 4×10^{-7} per frame. This corresponds to roughly 4×10^{-2} false alarms per hour, or one false alarm per 2.5 hours.

The false alarm rate will often be limited by clutter rather than by sensor noise. Clutter will result from clouds and terrain objects, appearing to have features similar to the target.

7.9 Validation Techniques

A short anecdotal detour is in order under this section heading. In engineering school, my professor followed a policy whereby a correct answer in the exam deserved a pass, 50%, but no more. His argument was that you need to be correct to pass; anything less is flunking. To obtain a distinction, your work must be distinctive, deserving the extra marks. In particular, you must demonstrate that you know the answer is correct. Validation of your answer gives you the right to rise above the rest, to present yourself in confidence in an elite group. If you validate, your work and your viewpoint are accepted.

Validation has little in common with radiometry, and yet it is an important step in radiometric analysis and modeling. See Section B.3 for a very brief introduction to this very important topic.

Bibliography

- [1] Lloyd, J. M., *Thermal Imaging Systems*, Plenum Press, New York (1975).
- [2] Hovanessian, S. A., *Introduction to Sensor Systems*, Artech House, Norwood, MA (1988).
- [3] Wittenstein, W., "Thermal range model TRM3," *Proc. SPIE* **3436**, 413–424 (1998) [doi: 10.1117/12.328038].
- [4] Vollmerhausen, R. H. and Jacobs, E., "The Targeting Task Performance (TTP) Metric A New Model for Predicting Target Acquisition Performance," Tech. Rep. AMSEL-NV-TR-230, NVESD, U.S. Army CERDEC, Fort Belvoir, VA 22060 (2004).

- [5] Helstrom, C. W., "Performance of Receivers with Linear Detectors," *IEEE Trans. Aerospace and Electronic Systems* 26, 210–217 (1990).
- [6] Hecht, E., *Optics*, 4th Ed., Addison Wesley, Boston, MA (2002).
- [7] Nicodemus, F. E., "Normalization in Radiometry," *Applied Optics* 12, 2960–2973 (1973).
- [8] Palmer, J. M. and Tamasco, M. G., "Broadband radiometry with spectrally selective detectors," *Optics Letters* 5(5), 208 (1980).
- [9] DeWitt, D., "Inferring Temperature from optical radiation measurements," *Optical Engineering* 25(4), 596–601 (1986) [doi: 10.1117/12.7973867].
- [10] Carmichael, G. W., "Issues in calibrating infrared seekers at low energy levels," *Proc. SPIE* 344, 34–42 (1982) [doi: 10.1117/12.933748].
- [11] Kaminski, W. R., "Range calculations for IR rangefinder and designators," *Proc. SPIE* 227, 65–79 (1980) [doi: 10.1117/12.958748].
- [12] Tomiyama, K., Pierluissi, J., and Hall, J. T., "Detection range computation by convolution for infrared detectors," *Proc. SPIE* 366, 157–164 (1982) [doi: 10.1117/12.934243].
- [13] Minkoff, J., *Signal Processing Fundamentals and Applications for Communications and Sensing Systems*, Artech House, Norwood, MA (2002).
- [14] Trishenkov, M. A., *Detection of Low-Level Optical Signals*, Kluwer Academic Publishers, Norwell, MA (1997).
- [15] Hippenstiel, R. D., *Detection Theory: Applications and Digital Signal Processing*, CRC Press, Boca Raton, FL (2002).
- [16] RCA Corporation, *RCA Electro-Optics Handbook*, no. 11 in EOH, Burle (1974).

Problems

- 7.1 Explain what the range equation is and why it is important. Derive a mathematical formulation for the range equation and elaborate on how it can be solved. [2]
- 7.2 Explain what happens when a small object is viewed in an image at different ranges. Draw a diagram that shows the pixel signal for different ranges from very close to very far. Divide the diagram into different regions, where different signal sources contribute to the pixel signal. Explain the dominant source in each region. [5]

- 7.3 Explain the term 'effective transmittance' and show how it is calculated. [2]
- 7.4 Provide a description (in words or equations, as applicable) of each of the following terms and explain where it is used: [10]
1. Signal-to-noise ratio (unitless).
 2. D^* - Dee-star, units $[\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}]$.
 3. Noise equivalent power - units $[\text{W}]$.
 4. Probability of detection (unitless).
 5. Point spread function (unitless).

Chapter 8

Optical Signatures

*It is the mark of an instructed mind
to rest satisfied with the degree of precision
which the nature of the subject permits
and not to seek an exactness
where only an approximation of the truth is possible.
Aristotle, Nicomachean Ethics*

8.1 Model for Optical Signatures

An optical signature is the manifestation of the radiometric characteristics of an object. The signature is formed by self-emitted flux, transmitted flux, and flux reflected from the object's surface. The magnitude of the different signature components depends on the state of the object itself (e.g., internal temperature) as well as the state of its environment (e.g., incident sunlight). The environment can also affect the long-term signature properties such as an increase of the object's temperature resulting from solar irradiance. Some objects' signatures may also depend on the internal state of the object (e.g., aircraft engine setting).

Optical signatures also have three-dimensional spatial properties. The object's intensity varies with view angle around the object. The calculation or measurement of the optical signature from one view is not always indicative of its signature from another view. Figure 8.1 shows three-dimensional spherical plots of calculated¹ contrast intensity signatures. The models used in these calculations are physically accurate models validated by measurement at selected view angles (Appendix B).

A conceptual description² of the main contributors to the apparent radiance from a small, *nominally uniform*, semi-transparent Lambertian surface with uniform surface temperature in open sunshine is shown in Equation (8.1), and Figure 8.2. Signatures for more-complex objects can be constructed as collections of signatures from such small areas.

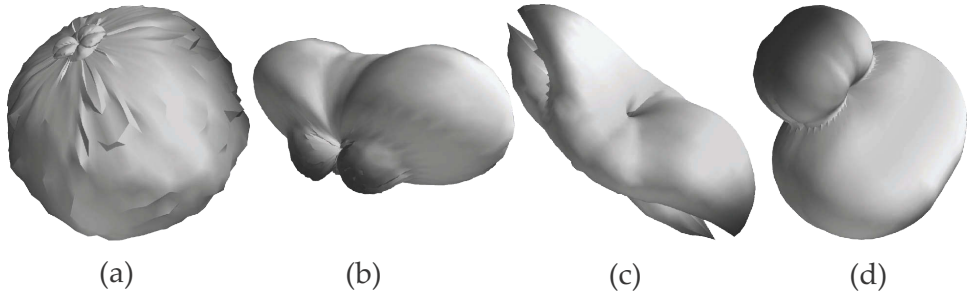


Figure 8.1 Spherical plots of aircraft contrast intensity signatures: (a) fighter aircraft in 3–5 μm , (b) helicopter in 8–12 μm , (c) fighter aircraft around 1 μm , and (d) transport aircraft in 8–12 μm .¹

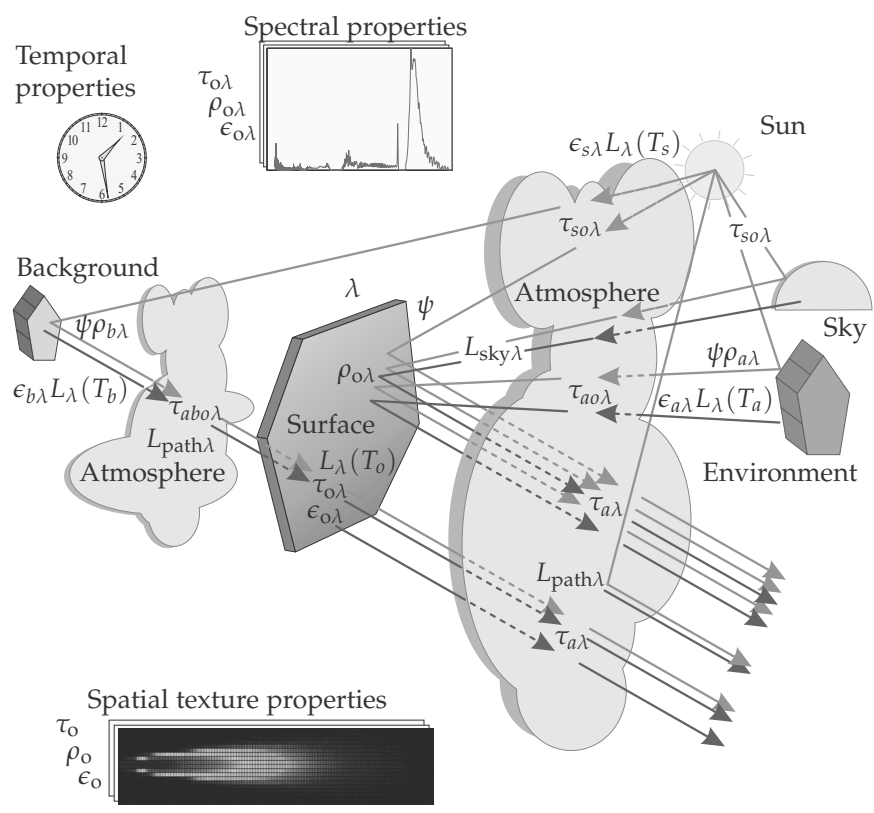


Figure 8.2 Main contributors to the radiometric signature.

In the real world it is impossible to fully know and describe an optical signature. But with a little effort, remarkably good approximations can be achieved. More information on signatures can be found in several sources.^{3–15}

The optical signature equation is not meant to be mathematically rigorous, but it serves to define the various signature contributions in a concise manner. The formulation in Equation (8.1) makes provision for surfaces with uniform radiance or for surfaces with small variations in radiance. Radiance variations are described by scaling ‘textures’ that modulate the integrated wideband radiance value. These textures are essentially small spatial variations in emissivity or reflectance. Virtually all of the elements in the equation are spectrally variant; spectral integrals provide the wideband radiance:

$$\begin{aligned}
 L_S = & \overbrace{\Delta_\epsilon \int_0^\infty \epsilon_{o\lambda}(\theta_v) L_\lambda(T_o) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{thermally emitted } L_{\text{self}}} \\
 & + \overbrace{\int_0^\infty L_{\text{path}\lambda} \mathcal{S}_\lambda d\lambda}^{\text{atmospheric path radiance } L_{\text{path}}} \\
 & + \overbrace{\int_0^\infty \tau_{o\lambda} \epsilon_{b\lambda} L_\lambda(T_b) \tau_{abo\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{transmitted background } L_{\text{trn back}}} \\
 & + \overbrace{\Delta_\rho \int_0^\infty \int_{\text{env}} \rho_{o\lambda} \epsilon_{a\lambda} L_\lambda(T_a) \tau_{ao\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected ambient background } L_{\text{ref amb}}} \\
 & + \overbrace{\Delta_\rho \cos \theta_a \int_0^\infty \int_{\text{sky}} \rho_{o\lambda} L_{\text{sky}\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected sky } L_{\text{ref sky}}} \\
 & + \overbrace{\Delta_\rho \psi \cos \theta_s \int_0^\infty f_r(\theta_s, \theta_v) \epsilon_{s\lambda} L_\lambda(T_s) \tau_{so\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{reflected sun } L_{\text{ref sun}}} , \quad (8.1)
 \end{aligned}$$

where the symbols are defined in Table 8.1.

All of the spectral integrals are weighted with the sensor’s spectral system response \mathcal{S}_λ . Consider the individual terms in Equation (8.1) as components in the signature:

1. **Self-emitted radiance (L_{self}):** The object emits flux according to its directional emissivity and Planck’s law (Section 3.5). For Lambertian surfaces the directional emissivity is simply the diffuse emissivity. The

Table 8.1 Terminology definition for Equation (8.1).

Symbol	Meaning
L_S	total radiance in the wavelength band S
$L_\lambda(T_s)$	spectral blackbody radiance, sun temperature T_s
$L_\lambda(T_a)$	spectral blackbody radiance, environment temperature T_a
$L_\lambda(T_b)$	spectral blackbody radiance, background temperature T_b
$L_\lambda(T_o)$	spectral blackbody radiance, uniform object temperature T_o
$L_{\text{path}\lambda}$	spectral atmospheric path radiance: emitted & scattered
$L_{\text{sky}\lambda}$	spectral sky radiance: emitted & scattered
$\epsilon_{s\lambda}$	solar surface's spectral emissivity
$\epsilon_{a\lambda}$	ambient environment's spectral emissivity
$\epsilon_{b\lambda}$	background spectral emissivity
$\epsilon_{o\lambda}(\theta_v)$	object surface directional spectral emissivity
Δ_ϵ	spatial texture variation in emissivity (unity if no texture)
$\rho_{o\lambda}$	object surface diffuse spectral reflection
Δ_ρ	spatial texture variation in reflectivity (unity if no texture)
$f_r(\theta_s, \theta_v)$	object surface bidirectional reflection distribution function
$\tau_{o\lambda}$	object spectral transmittance
$\tau_{a\lambda}$	object-to-sensor spectral atmospheric transmittance
$\tau_{abo\lambda}$	background-to-object spectral atmospheric transmittance
$\tau_{ao\lambda}$	ambient-to-object spectral atmospheric transmittance
$\tau_{so\lambda}$	sun-to-object spectral atmospheric transmittance
ψ	$A_{\text{sun}} / (d_{\text{sun}}^2 \pi) = 2.17 \times 10^{-5}$
A_{sun}	area of the sun
d_{sun}	distance to the sun
θ_a	angle between the surface normal and the vertical
θ_s	angle between the surface normal and solar incidence
θ_v	angle between the surface normal and the view direction
S_λ	measurement instrument spectral response

object's radiance is weighted by the atmospheric transmittance between the object and the sensor.

2. **Atmospheric path radiance (L_{path}):** The atmospheric path between the surface and the sensor adds radiance as described in Section 4.6.5.
3. **Transmitted background radiance ($L_{\text{trn back}}$):** Semi-transparent surfaces transmit flux from behind the surface. The background radiance, surface transmittance, and the atmospheric transmittance between the background and the surface determine flux exitant from the surface. For opaque surfaces this contribution is zero.
4. **Reflected ambient radiance ($L_{\text{ref amb}}$):** It is assumed that the flux from the environment is incident from all directions; the object is fully enclosed by the environment. This assumption is approximately valid when the object is indoors, but less so if the object is outdoors. However, even outdoors, the object may be immersed in the atmosphere, which also provides at least some environmental flux (Section 4.6.7 and Figure 4.12).
5. **Reflected sky radiance ($L_{\text{ref sky}}$):** Skylight is a diffuse source, caused by scattering and emission in the atmosphere. It is difficult to model accurately and is best measured or calculated with an atmospheric code (Section 4.6.5 and Figure 4.12). Rayleigh scatter (in the visual spectral range) can be spectrally approximated by a 1×10^4 -K source with low emissivity.
6. **Reflected solar radiance ($L_{\text{ref sun}}$):** A simple model for reflected sunlight is presented in Section 3.7. The reflected solar radiance depends on the orientation of the surface, the BRDF (f_r), and the transmittance of the atmosphere from the sun to the surface and then from the surface to the sensor.

8.2 General Notes on Signatures

The object's surface emissivity ϵ scales thermal exitance as well as the reflection from opaque surfaces by $\rho = 1 - \epsilon$. Surfaces with high emissivity are poor reflectors, and vice versa. Some objects' signatures are as much affected by emissivity variations as by temperature variations. There is no way to tell the difference between reflected flux or self-emitted flux in a measured signature.

Natural ground-object temperatures are mostly in the range -10 to $+100$ °C. The peak thermal exitance of these targets lies in the $8\text{--}12\text{-}\mu\text{m}$

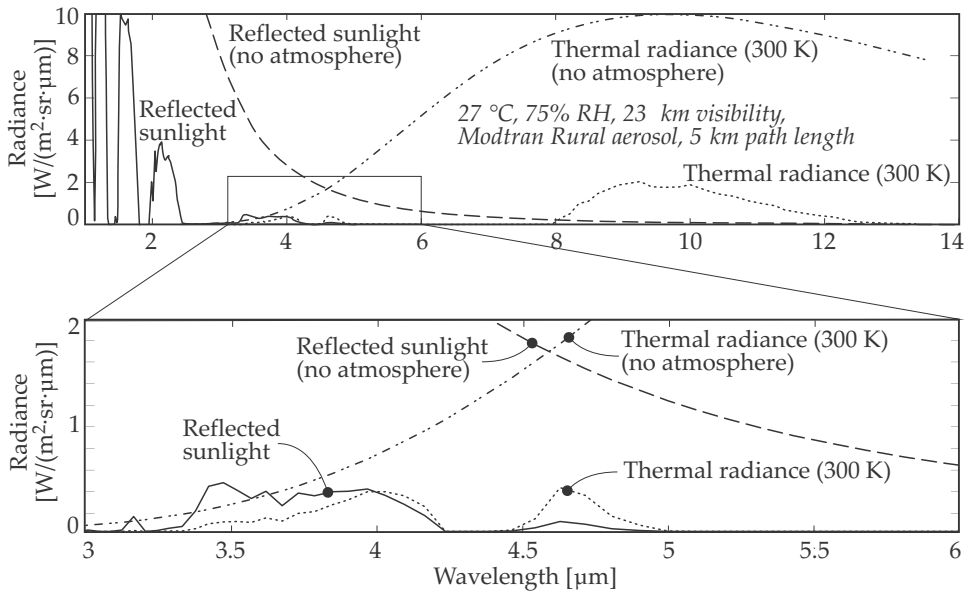


Figure 8.3 Reflected and emitted signature from a target with 80% emissivity.

spectral range. These same objects radiate a substantial signature in the 3–5- μm spectral band as well but very little in the visual spectral band. The practical signature of concern is often the target contrast (the difference between the target and its surroundings) rather than the absolute signature.

The relative contributions between reflected sunlight and self-emitted infrared flux are shown in Figure 8.3. The object is viewed through a tropical atmosphere (27 °C, 75% relative humidity) with 23-km visibility. The solid line represents radiation from a 300-K earth-bound object, whereas the dotted line shows the sunlight reflected from the same surface. Note that in the 8–12- μm band, the self-emitted flux dominates the signature, whereas in the shorter-wavelength bands, the reflected sunlight dominates the signature. The 3–5- μm spectral band shows equal contributions from self-emitted and reflected sunlight for objects at ambient terrain temperature. Figure 8.3 indicates that in the 3–5- μm spectral band there is a significant infrared signal in the absence of sunlight. There is also a significant contribution from reflected sunlight.

8.3 Reflection Signatures

The Phong Equation (3.39) provides a simple (but not very accurate) model for BRDF:

$$f_{r,\text{Phong}} = \frac{\rho_o}{\pi} + \frac{\rho_s(n+1)\cos^n\alpha}{2\pi\cos\theta_i}.$$

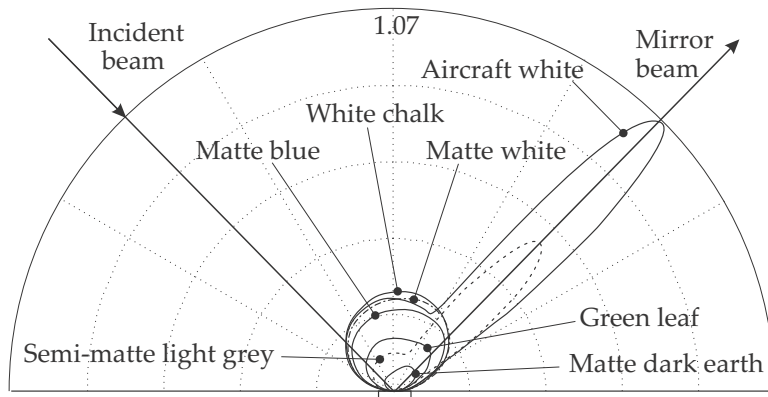


Figure 8.4 Phong BRDF for paints and natural surfaces in the NIR spectral band.

Table 8.2 Phong BRDF parameters for paints and natural surfaces in the NIR spectral band.

Surface finish	0.75–1.4- μm NIR band		
	ρ_o	ρ_s	n
Matte dark earth paint	0.11	0.033	6
Matte blue paint	0.52	0.022	8
Matte white paint	0.58	0.03	6
Semi-matte light grey paint	0.25	0.052	44
Aircraft white paint	0.6	0.05	80
Natural soil	0.33		
Sea sand	0.49		
Green leaf	0.43	0.015	15
Gypsum / white chalk	0.77	0.044	3
Steel plate, freshly coated with Zn	0.05	0.39	160
Concrete	0.55		

It is, however, adequate for the purpose of this book — mathematically simple, yet sufficiently illustrative. Measurements on several paints and natural surfaces¹⁶ indicated BRDF values as shown in Table 8.2 and Figure 8.4. The application of this data toward a laser rangefinder performance calculation is presented in Section 9.4.8.

8.4 Modeling Thermal Radiators

Planck's law is not always a good approximation for real-world objects. Such objects are better modeled as a grey body or spectrally selective thermal radiators. The method described here models the source as a black-body thermal radiator (at a specific temperature) multiplied with a spectral

emissivity and an effective area. This approach applies only to thermal sources complying with Planck's law. The procedure assumes a single, uniform, radiating element. Complex target objects must be modeled as a collection of uniform radiating elements.

Equation (6.13) describes the flux flowing from the radiator to the receiver by an equation of the form (simplified here to retain only the essential information):

$$\Phi = \frac{A_0 A_1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda}(T_{bb}) \tau_{m\lambda} d\lambda, \quad (8.2)$$

where $L_{0\lambda}(T_{bb})$ is a Planck-law radiator at temperature T_{bb} , and the transmittance factor $\tau_{m\lambda} = \tau_{a\lambda} \tau_{s\lambda} \alpha_{s\lambda}$ collects the various transmittance factors as a single variable. The modeling objective is to find A_0 , T_{bb} , and $\epsilon_{0\lambda}$, given all of the other (measured and given) information. These three parameters are sufficient to characterize and model any thermal radiator.

Signature modeling requires information and knowledge beyond just the measured data. There must be some indication of the temperature or area (however vague). Emissivity is bounded by zero and one, but even within this range, some information should be available. The additional information is required to (a) guide the model analysis and (b) provide partial validation for the model results.

Equation (8.2) is a mathematical statement and many different combinations of A_0 , T_{bb} , and $\epsilon_{0\lambda}$ provide equally valid *mathematical* solutions to the equation. For example, for different values of source temperature, corresponding values of spectral emissivity can be calculated, all of which are valid in a mathematical sense. However, modeling requires physically viable solutions, not just mathematically valid solutions. It follows that only a few mathematical solutions are physically viable solutions, and of these, only one is a valid physical solution. Section 9.8 investigates this approach in a practical example.

Extrapolation is generally not safe, both in a mathematical and physical sense. There is, however, an implied requirement to perhaps slightly extrapolate around the calculated model parameters. If the model parameters are a valid physical solution, limited extrapolation should be in order. If, however, the model parameters do not represent a valid physical solution, any extrapolation is in error.

Some radiators (gas plumes and flames) have a nebulous nature; the area depends on the radiance levels considered. If only the most intense part of the flame is measured, the area is very small, whereas the area can be large if all low-radiance regions are also included. This problem is investigated in Section 8.4.2.

Practical instruments measure and represent the optical flux as a voltage or digital count, effectively implementing Equation (6.16). A process known as calibration⁵ is used to determine the relationship between the measured voltage and the flux on the detector $\Phi_S = f(v_S)$ in the band \mathcal{S} . This relationship is known as the instrument function.

With the preliminaries in place, the modeling effort should proceed along these steps:

1. Estimate a mathematically valid spectral emissivity.
2. Using the estimated spectral emissivity, calculate [using Equation (8.2)] any appropriate combination of temperature and area that would yield the measured results.
3. Select/confirm the area and temperature by measuring at least one of the two parameters in some other way.
4. Review the combination of spectral emissivity, temperature, and area for physical validity (or at least physical viability). Repeat the process if one of the parameters is not valid.

This process entails an element of uncertainty and guesswork. The uncertainty in itself does not invalidate the process; it is still a useful and valid process. However, accurate results require that assumptions be verified and model parameters validated.

8.4.1 Emissivity estimation

A spectroradiometer can be used to measure the spectral radiance of the target source. After conversion by the instrument function, the radiometer provides radiance measurement according to Equation (6.16), but for a very narrow spectral width $\lambda_2 - \lambda_1 = \delta\lambda \rightarrow 0$,

$$L_{m\lambda} = \frac{k \hat{\mathcal{R}} Z_t dA_0 \cos \theta_0 A_1 \epsilon_{0\lambda} L_{0\lambda}(T_{bb}) \tau_{a\lambda} \mathcal{S}_\lambda}{R_{01}^2} \delta\lambda, \quad (8.3)$$

where $L_{m\lambda}$ is the measured radiance, and k is the instrument calibration function (assumed here to be a simple gain factor). The instrument function calibration assumes that the radiator fills the complete radiometer FOV. This assumption is not always met, in which case the calculation must be adapted to account for the instrument's response to a partially filled FOV. If the target source has much higher radiance than the background, the background can be ignored. In this case, the spectral emissivity shape will be correct, but its absolute scaling will be unknown.

In order to determine the spectral emissivity, Equation (8.3) must be inverted to solve for $\epsilon_{0\lambda}$. This cannot be done analytically but is easily achieved by numerical methods. Because the spectroradiometer measures the *spectral radiance* $L_{\lambda m}$ directly, the spectral emissivity can be determined by solving

$$L_{m\lambda} = \epsilon_{0\lambda} L_{0\lambda}(T_{bb}) \quad (8.4)$$

by dividing both sides by the blackbody radiance

$$\epsilon_{0\lambda} = \frac{L_{m\lambda}}{L_{0\lambda}(T_{bb})} . \quad (8.5)$$

The key step is to select a source temperature T_{bb} that yields reasonable emissivity values. From the definition of emissivity, emissivity may not exceed unity or be negative. Note that, within physically viable limits, any choice of peak emissivity combined with temperature can be made. There is usually some a priori knowledge of the spectral emissivity or temperature that would forward a preferred solution. See Section 9.8 for a practical application of this principle.

Solids normally have an expected range of emissivity values. Also, over a relatively narrow spectral range, the infrared emissivity for a solid object is expected to be more or less constant — this observation may help in estimating the temperature. If the source target is an opaque material known to have near-constant spectral emissivity, the physically valid temperature is the value that yields a near-constant spectral emissivity.

If there are large quantities of radiating species in a gas, it may be assumed that the gas is optically thick at some wavelength. An optically thick gas radiator would result in an emissivity near unity at some wavelength. The objective is then to determine the temperature that yields the appropriate maximum emissivity value. Open yellow flames may have emissivity values between 0.0 and 0.2, (or even higher) depending on the concentration of solids (e.g., carbon soot) in the flame.

The calculated emissivity and its associated temperature are taken as the first two model parameters. The closer any one of these two parameters are to the real physical situation, the more accurate the model will be in physical realism.

8.4.2 Area estimation

If the radiating object is a uniformly radiating solid, it is relatively easy to determine the area. However, if the solid body does not radiate uniformly over its surface, or if the radiator is a gas, the radiating area [in terms of Equation (8.2)], is not well defined.

The magnitude of ‘radiator area’ depends on the exact definition of what is meant by ‘area.’ If a thermal image of the source is available, three radiator area definitions are possible:

1. Set a radiance threshold at some level and ignore all pixels in the image with values below this threshold. The problem with this approach is that the selected pixels may have widely varying radiance values, from very cool to very hot. A *more accurate* approach is to segment the object image into several regions, each with a limited range of pixel radiance levels (somewhat similar to a terrain contour map).
2. Set a spatial limit to the size of the radiator and ignore all pixels outside the selected size. This approach may reject a significant amount of radiation. However, if the spatial solid angle is selected carefully, this approach may yield a good estimate of the useful radiator area.
3. Integrate the flux over a large spatial extent and at all radiance levels — use as large an area as feasible. Peak normalize (Section 7.2.3) the measured flux by assuming the highest radiance level for the whole radiator area, and then calculate an ‘area’ to provide the same flux as the integrated measured flux.

Using the spectral emissivity calculated in the previous section, sets of effective areas and temperatures can be calculated that would yield the required measured flux. In other words, given $\epsilon_{0\lambda}$, find A_1 and T_{bb} that would solve Equation (8.2) for the measured data. By investigating the balance between temperature and area, one may develop a better understanding of the radiating nature of the target object.

Figure 8.5 shows 3–5- μm radiance images for a flame measurement obtained with a Cedip MWIR thermal camera. The left figure shows the original image. The six images on the right show the flame at increasing threshold levels, from a low value (just above the background) to a value near the peak of the image. These images were obtained by setting threshold levels to exclude lower pixel values from the image (called greylevel segmentation of the image). The threshold radiance value and size of the segmented part of the image is displayed with each picture. The code for this analysis is shown in Section D.5.5.

The flame size was calculated using the simple threshold segmentation method and also the peak normalized method. The results are shown in Figure 8.6. Note that the peak normalized area is almost five times smaller than the simply threshold-segmented area at the lowest threshold value.

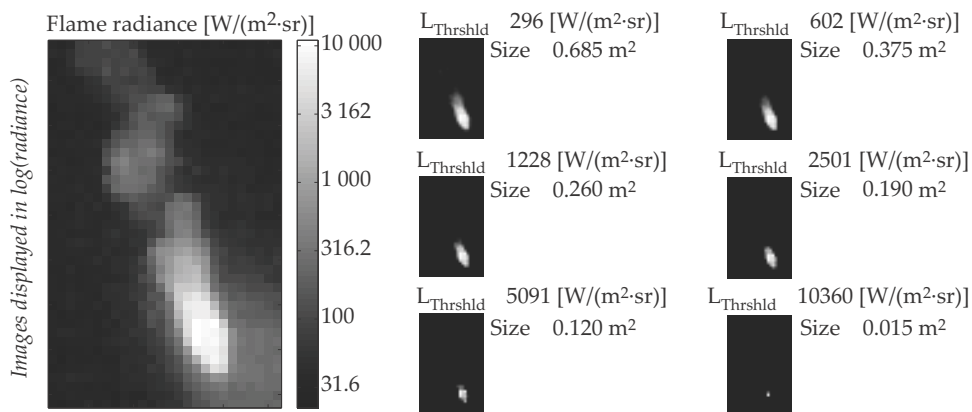


Figure 8.5 Measured flame radiance in the 3–5- μm spectral band for six different threshold levels.

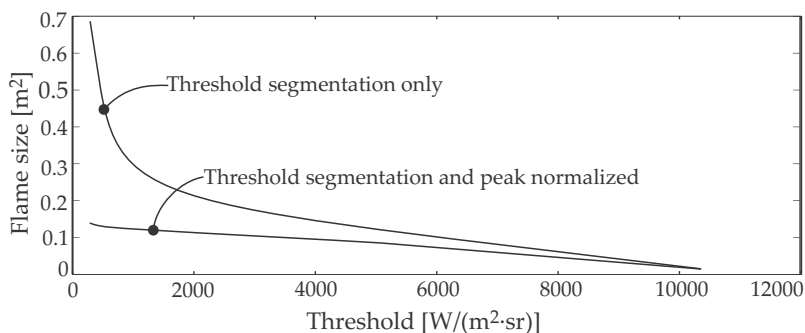


Figure 8.6 Flame-size predictions as a function of segmentation threshold.

Inspection of the segmented pictures in Figure 8.5 might lead the analyst to decide that the flame area is approximately 0.2 m^2 , corresponding to a threshold of $2500 \text{ W}/(\text{m}^2 \cdot \text{sr})$. In this case the area is selected on the basis of the analyst's evaluation of the threshold-segmented pictures.

The analysis indicates that the flame has large areas of relative low radiance, which contributes relatively little to the total signature. Depending on the application, different flame areas can be chosen. If shape is not important, the flame could be modeled as a simple uniform radiator with area of 0.14 m^2 . If shape is important, the radiance gradient across the flame must be accounted for, and the model becomes more complex.

8.4.3 Temperature estimation

Temperature measurement is a nontrivial measurement activity.^{17,18} An object may even possess different temperatures, depending on the process

or instrument used to measure the temperature. When building signature models, temperature estimation can be assisted by setting constraints on the temperature range. Systematic narrowing of these constraints will result in a band of conceivable temperatures. Prior experience, intensive literature searches, or expert guidance is essential in this process. This section describes two techniques that require access to a spectral radiance measurement of the test sample.

When estimating the temperature, emissivity presents a constraint by its unity maximum value. If $L_{\lambda m}$ is a measured spectral radiance, and $L_{bb\lambda}(T_m)$ is the blackbody radiance at temperature T_m , the following equation sets the lower temperature constraint:

$$\epsilon_{\lambda m} = \frac{L_{m\lambda}}{L_{bb\lambda}(T_m)} \leq 1. \quad (8.6)$$

Figure 9.19 shows a case where a too-low temperature estimate requires an emissivity exceeding unity, which is impossible.

If an object is known to have a near-constant spectral emissivity, the object's temperature can be roughly estimated by matching the object's spectral radiance to a scaled version of the Planck-law equation. In this case, the scale factor will be the emissivity of the object. Figure 8.7 shows the spectral radiance of two solid surfaces, both measured and calculated. The measured radiance values were not corrected for atmospheric transmittance. Superimposed on the measured data are shown several calculated Planck-law curves. The curve that fit best is an indication of the temperature of the surface. This approach is sensitive to spectral atmospheric transmittance effects and any variation in emissivity of the sample. In Figure 8.7(a), the calculated radiance values were scaled to the measured radiance value at $3.95 \mu\text{m}$. In Figure 8.7(b), all of the calculated values were scaled by 0.85, the estimated emissivity of the plate. This method was used to determine the temperature of a carbon-rich flame [Magnesium-Teflon[®]-Viton[®] (MTV) flare] with reasonably good success.

Two-color or multi-color temperature measurement¹⁷ is a variation of the technique described above. The object's radiance is measured in two (or more) spectral bands, and the ratio between the bands is an indication of the temperature. The two spectral bands must be selected to provide sufficient variation in color ratio over the desired temperature range. Two-color ratio measurement is used successfully to discriminate between aircraft and MTV flares in two-color missile seekers. In this application the objective is not accurate temperature measurement but rather to provide a threshold for decision making.

Temperature estimation depends heavily on the use of reference or

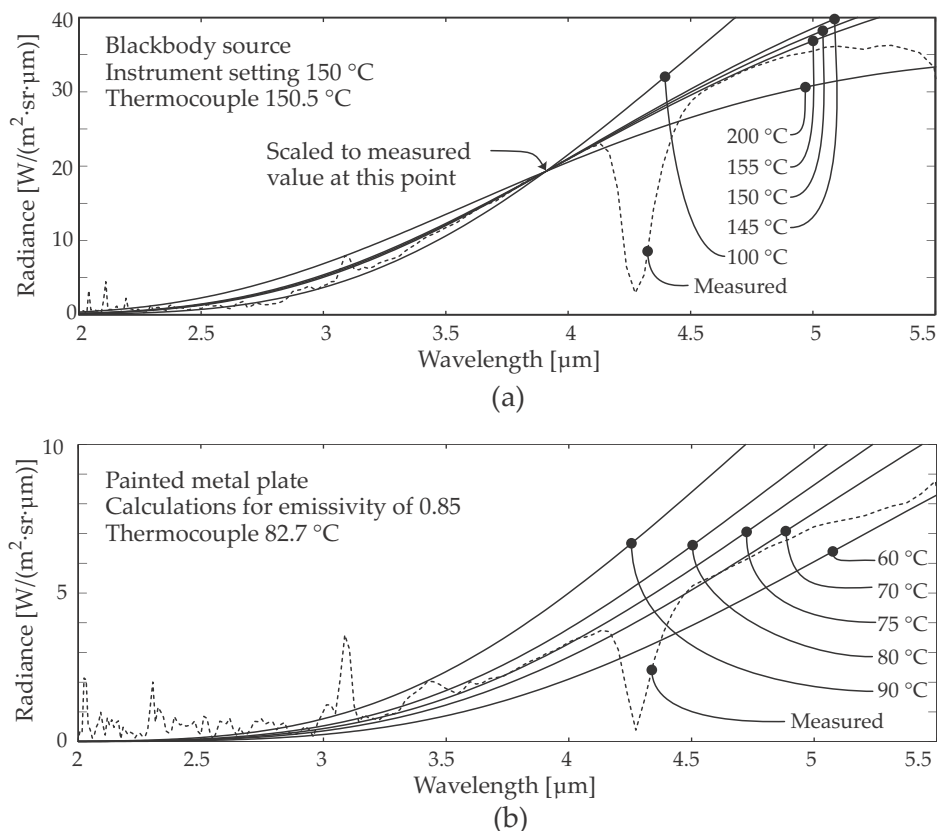


Figure 8.7 Comparison of measured and calculated radiance for solid objects: (a) blackbody and (b) painted metal plate.

external information.¹⁹ Chemists can often calculate or predict the temperature of burning chemical compounds.²⁰ Internet search results should be used with caution, especially with regards to information on the temperature of flames.

The effect of surface emissivity on apparent temperature is discussed in Section 3.2.4.

8.5 Measurement Data Analysis

'Same as' measurements are commonly employed in infrared measurements. The object radiance is compared with the radiance measured from a known calibration source. The process can be best described as follows: *The object was observed at long range, and the measured irradiance was comparable to the irradiance observed when viewing a known calibration source at a short range.* The following information is typically available: the radiometer

FOV, spectral response, and the meteorological data during the measurement. This procedure also requires that some estimate of the object spectral emissivity be known (either by measurement or research investigation).

The purpose of this data analysis calculation is to determine the object temperature, given the object spectral emissivity, area, and atmospheric conditions at the time of the measurement. The intention is to calculate the temperature directly from the measured data with no intermediate steps. This approach requires a mathematical formulation that can be solved numerically.

Note that the temperature so determined is not necessarily a physically valid temperature. It is the value mathematically required by the chosen area and emissivity to provide the measured signal. If the area and emissivity values are physically valid, then the temperature is also a physically valid temperature.

Equation (6.16) describes the signal voltage at the output of the radiometer and is simplified to

$$v_S = \frac{k \hat{R} Z_t A_0 A_1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} S_\lambda d\lambda. \quad (8.7)$$

The test results are reported in terms of two measurements giving the same voltage in the instrument. The first voltage is the calibration voltage when observing a known source. The second voltage is the measurement of the target source.

Assuming that the calibration is done with a blackbody source with emissivity close to unity, over a very short distance, the voltage for the calibration measurement is

$$v_c = k \hat{R} Z_t \Omega_p A_1 \int_0^\infty L_{0\lambda}(T_c) S_\lambda d\lambda, \quad (8.8)$$

where the instrument FOV Ω_p is filled by the calibration source, T_c is the blackbody temperature, and v_c is the instrument output voltage during this calibration.

From Equations (8.1) and (8.7), the voltage for the object measurement is

$$v_S = \frac{k \hat{R} Z_t A_0 A_1}{R_{01}^2} \underbrace{\Delta_\epsilon \int_0^\infty \epsilon_{o\lambda}(\theta_v) L_\lambda(T_o) \tau_{a\lambda} S_\lambda d\lambda}_{\text{thermally emitted } L_{\text{self}}} + k \hat{R} Z_t \Omega_p A_1 \underbrace{\int_0^\infty L_{\text{path}\lambda} S_\lambda d\lambda}_{\text{atmospheric path radiance } L_{\text{path}}}$$

$$\begin{aligned}
& + \frac{k \widehat{\mathcal{R}} Z_t A_0 A_1}{R_{01}^2} \overbrace{\int_0^\infty \tau_{o\lambda} \epsilon_{b\lambda} L_\lambda(T_b) \tau_{abo\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{transmitted background } L_{\text{trn back}}} \\
& + \frac{k \widehat{\mathcal{R}} Z_t A_0 A_1}{R_{01}^2} \Delta_\rho \overbrace{\int_0^\infty \int_{\text{env}} \rho_{o\lambda} \epsilon_{a\lambda} L_\lambda(T_a) \tau_{ao\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected ambient background } L_{\text{ref amb}}} \\
& + \frac{k \widehat{\mathcal{R}} Z_t A_0 A_1}{R_{01}^2} \Delta_\rho \cos \theta_a \overbrace{\int_0^\infty \int_{\text{sky}} \rho_{o\lambda} L_{\text{sky}\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected sky } L_{\text{ref sky}}} \\
& + \frac{k \widehat{\mathcal{R}} Z_t A_0 A_1}{R_{01}^2} \Delta_\rho \psi \cos \theta_s \overbrace{\int_0^\infty f_r(\theta_s, \theta_v) \epsilon_{s\lambda} L_\lambda(T_s) \tau_{so\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{reflected sun } L_{\text{ref sun}}}.
\end{aligned} \tag{8.9}$$

Because the radiance during the measurement is the same as the radiance during calibration, now equate the values. Furthermore, if the target area exceeds the size of the pixel footprint (i.e., an extended source), A_s/R_{01}^2 becomes the instrument FOV Ω_p , and then the equation can be simplified considerably because the geometric factors are the same for all of the terms. Assuming that all of the other parameters are known, the target source temperature can now be obtained by solving for T_o in

$$\begin{aligned}
0 = & - \overbrace{\int_0^\infty L_{0\lambda}(T_c) \mathcal{S}_\lambda d\lambda}^{\text{calibration radiance}} \\
& + \overbrace{\Delta_\epsilon \int_0^\infty \epsilon_{o\lambda}(\theta_v) L_\lambda(T_o) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{thermally emitted } L_{\text{self}}} + \overbrace{\int_0^\infty L_{\text{path}\lambda} \mathcal{S}_\lambda d\lambda}^{\text{atmospheric path radiance } L_{\text{path}}} \\
& + \overbrace{\int_0^\infty \tau_{o\lambda} \epsilon_{b\lambda} L_\lambda(T_b) \tau_{abo\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{transmitted background } L_{\text{trn back}}} \\
& + \Delta_\rho \overbrace{\int_0^\infty \int_{\text{env}} \rho_{o\lambda} \epsilon_{a\lambda} L_\lambda(T_a) \tau_{ao\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected ambient background } L_{\text{ref amb}}} \\
& + \Delta_\rho \cos \theta_a \overbrace{\int_0^\infty \int_{\text{sky}} \rho_{o\lambda} L_{\text{sky}\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected sky } L_{\text{ref sky}}} \\
& + \Delta_\rho \psi \cos \theta_s \overbrace{\int_0^\infty f_r(\theta_s, \theta_v) \epsilon_{s\lambda} L_\lambda(T_s) \tau_{so\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{reflected sun } L_{\text{ref sun}}}.
\end{aligned} \tag{8.10}$$

Solving this form of the equation is challenge because of the large number

of parameters that must be known. In practice, several simplifications are commonly made, as shown in the next few sections.

8.6 Case Study: High-Temperature Flame Measurement

Consider the case where the radiometer is observing a flame. The flame temperature is much higher than the surrounding background and the atmosphere. It is known that flames have high transmittance and low emissivity. Suppose further that path radiance can be ignored. Several terms in Equation (8.10) are therefore discarded. The problem can then be stated as follows:

$$\overbrace{\Omega_p \int_0^\infty L_{0\lambda}(T_c) \mathcal{S}_\lambda d\lambda}^{\text{calibration radiance}} = \overbrace{\Delta_\epsilon \frac{A_0}{R_{01}^2} \int_0^\infty \epsilon_{o\lambda}(\theta_v) L_\lambda(T_o) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{thermally emitted } L_{\text{self}}}, \quad (8.11)$$

where the requirement for an extended target size has been removed. A smaller target size results in a target solid angle smaller than the instrument solid angle. This may occur during practical measurements where flames are measured over long range. The solution to Equation (8.11) can be readily obtained by numerical analysis, but it does require a known spectral emissivity (see Section 8.4.1).

Two important observations can be made on the methodology given here: (1) the atmospheric transmittance is compensated for by multiplication (no divide-by-zero errors)- and (2) the absolute value of the radiometer system response does not appear in the solution; only the spectral system response \mathcal{S}_λ is required.

8.7 Case Study: Low-Emissivity Surface Measurement

The emissivity of shiny metallic surfaces are lower than the emissivity of rough Lambertian surfaces. The effect of this low emissivity on temperature measurements should be carefully considered. Suppose the irradiance from a metal surface is measured over a short distance, as an extended source, with the intention to determine the temperature of the surface. For this analysis start with Equation (8.10), assume an extended target, and ignore the terms for path radiance, transmitted background flux, reflected sky, and reflected sunlight. Keep the terms for self-exittance and terrain background. The terrain background in this case is meant to model an enclosed volume, such as in the laboratory. Note that for an opaque surface, $\rho_{o\lambda} = (1 - \epsilon_{o\lambda})$. This means that as the self-radiance decreases due to lower emissivity, the reflected ambient radiance increases due to reflection.

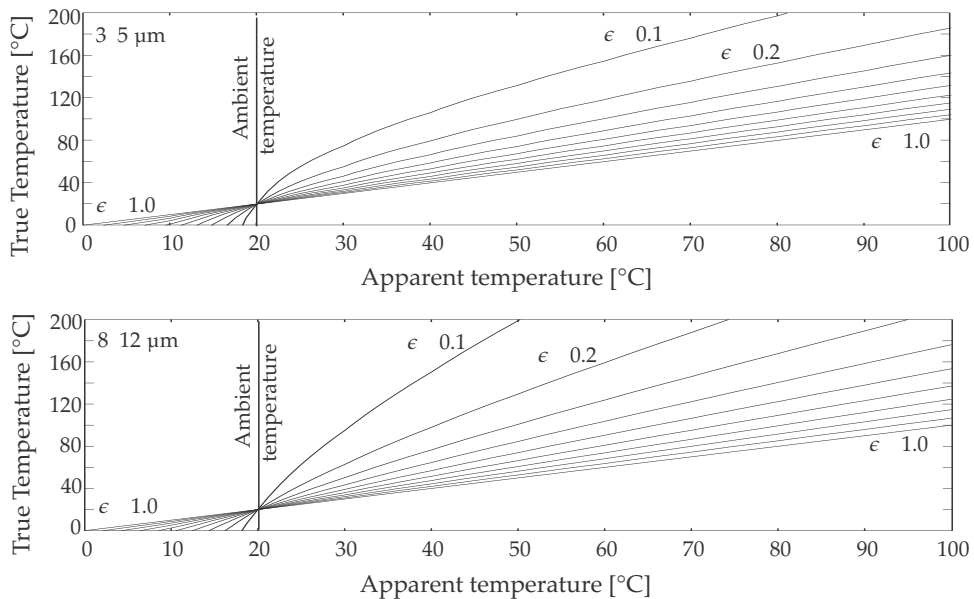


Figure 8.8 Apparent versus real temperature for different source emissivity values, ranging from 1 (straight line) down to 0.1 (most curved line) in steps of -0.1 , in the $3\text{--}5\text{-}\mu\text{m}$ and $8\text{--}12\text{-}\mu\text{m}$ spectral bands.

The source temperature can now be determined by solving

$$\begin{aligned}
 \overbrace{\int_0^\infty L_{0\lambda}(T_c) \mathcal{S}_\lambda d\lambda}^{\text{measured radiance}} &= \overbrace{\int_0^\infty \epsilon_{o\lambda} L_\lambda(T_o) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{thermally emitted } L_{\text{self}}} \\
 &\quad + \underbrace{\int_0^\infty (1 - \epsilon_{o\lambda}) \epsilon_{a\lambda} L_\lambda(T_a) \tau_{ao\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}_{\text{diffuse reflected ambient background } L_{\text{ref amb}}}. \quad (8.12)
 \end{aligned}$$

Equation (8.12) is now applied in a somewhat different manner. The right side is an accurate calculation of the signature as a function of target emissivity. Suppose the left side represents a noncontact temperature measurement instrument, but calibrated to an assumed target emissivity of one. Then T_c will represent the measured temperature of the test target. This apparent temperature T_c will be incorrect because of the incorrectly assumed unity emissivity. The physical temperature versus the erroneous observed temperature, as a function of object emissivity, is shown in Figure 8.8.

8.8 Case Study: Cloud Modeling

The objective with this case study is to derive a first-order empirical model for the radiance in the ‘silver lining’ edge of a back-lit cloud. The problem is solved by analyzing measured data and compiling a simple model. In this instance the only available measured information was a ‘temperature’ measurement in imaging-camera images. No other infrared or calibration information was available. The *modus operandi* was to first calculate the radiance from the measured temperature data and second to build a model around the calculated radiance values.

8.8.1 Measurements

Measurements were made at varying angles from the sun, ranging from 10 deg to 90 deg from the sun. The distance was measured with a laser rangefinder and varied from one to three kilometers. The measurement was performed from ground-level at 1500 m above sea level. The ground level ambient temperature during the measurements was 20 to 25 °C. In all of the subsequent calculations the LOWTRAN (precursor to MODTRAN™) Mid-latitude Summer model was used. The atmospheric transmittance and path radiance values are shown in Figures 8.9(b) and 8.9(c).

Measurements were performed with an InSb MWIR camera fitted with an antisolar filter. This filter suppresses all radiation at wavelengths shorter than 3 μm. The camera was set to assume a target emissivity of one. The vendor-supplied software was used to read temperature values directly from the image, so this analysis starts with a measured temperature value as input. The camera instrument function was calibrated against a blackbody simulator at close range under laboratory conditions.

The measurements indicated that a typical side-lit cloud has an apparent temperature of between 10 and 20 °C (Table 8.3). Back-lit clouds had silver linings and small hot spots with apparent temperatures between 30 °C and 50 °C. This increase in apparent temperature is due to forward scattering by the silver lining — not a high cloud temperature. The first step is to calculate the cloud radiance values from the recorded temperatures. The cloud radiance is calculated using the calibration equation

$$L_{\text{cloud}} = \int_0^{\infty} \epsilon_{0\lambda} L_{0\lambda}(T_m) S_{\lambda} \tau_{m\lambda} d\lambda, \quad (8.13)$$

where $\epsilon_{0\lambda} = 1$ is the calibration source emissivity, T_m is the cloud temperature as indicated by the camera, S_{λ} is the normalized camera response shown in Figure 8.9(a), and $\tau_{m\lambda}$ is the atmospheric transmittance during calibration.

Table 8.3 Measured and modeled cloud data.

Cloud sample	Apparent temperature	Measured radiance	Silver-lining factor ζ		Model radiance
			$T_c = 10^\circ\text{C}$	$T_c = 0^\circ\text{C}$	
Side-lit	10	1.34	1	1	1.16
Back-lit	30	2.7	20	23	2.7
Back-lit	50	5	50	51	5
Path		0.53			
	$^\circ\text{C}$	$\text{W}/(\text{m}^2\cdot\text{sr})$		-	$\text{W}/(\text{m}^2\cdot\text{sr})$

8.8.2 Model

In order to describe the forward scattering in the cloud's silver lining, define a scaling factor ζ called the silver-lining factor. The silver-lining factor is the ratio by which a back-lit cloud transmits (scatters) more than a side-lit cloud would reflect. It is calculated on the assumption that the spectral signature is the same for reflected and forward-scattered light. The silver-lining factor value is never less than unity. A cloud with no forward scattering, i.e., a side-lit cloud, has a silver-lining value of unity. As the forward scattering increases, silver-lining scatter also increases. Note that the silver-lining factor is not directly based on any physical theory; it is an empirical model.

Start from Equation (8.1), and remove all of the terms except those for self-radiance, path radiance, sky radiance, and reflected sunlight. The solar incidence angle $\cos \theta_s = 1$ and the object normal angle $\cos \theta_a = 1$ because the cloud is an extended target. Furthermore, the cloud is a diffuse spectral reflector; hence $f_r(\theta_s, \theta_v) = \rho_{o\lambda}$, and $\epsilon_{o\lambda}(\theta_v)$ is a diffuse spectral emissivity. The cloud is opaque, hence $\rho_{o\lambda} = (1 - \epsilon_{o\lambda})$. The cloud radiance model is then

$$\begin{aligned}
 L_S = & \overbrace{\int_0^\infty (1 - \rho_{o\lambda}) L_\lambda(T_o) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{thermally emitted } L_{\text{self}}} + \overbrace{\zeta \psi \int_0^\infty \rho_{o\lambda} \epsilon_{s\lambda} L_\lambda(T_s) \tau_{so\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{reflected sun } L_{\text{ref sun}}} \\
 & + \overbrace{\int_0^\infty L_{\text{path}\lambda} \mathcal{S}_\lambda d\lambda}^{\text{path radiance } L_{\text{path}}} + \overbrace{\int_0^\infty \int_{\text{sky}} \rho_{o\lambda} L_{\text{sky}\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\Omega d\lambda}^{\text{diffuse reflected sky } L_{\text{ref sky}}}, \quad (8.14)
 \end{aligned}$$

where T_o is the cloud temperature, and the remaining parameters are as previously defined. The spectral cloud reflectance $\rho_{o\lambda}$ is shown in Figure 8.9(a).

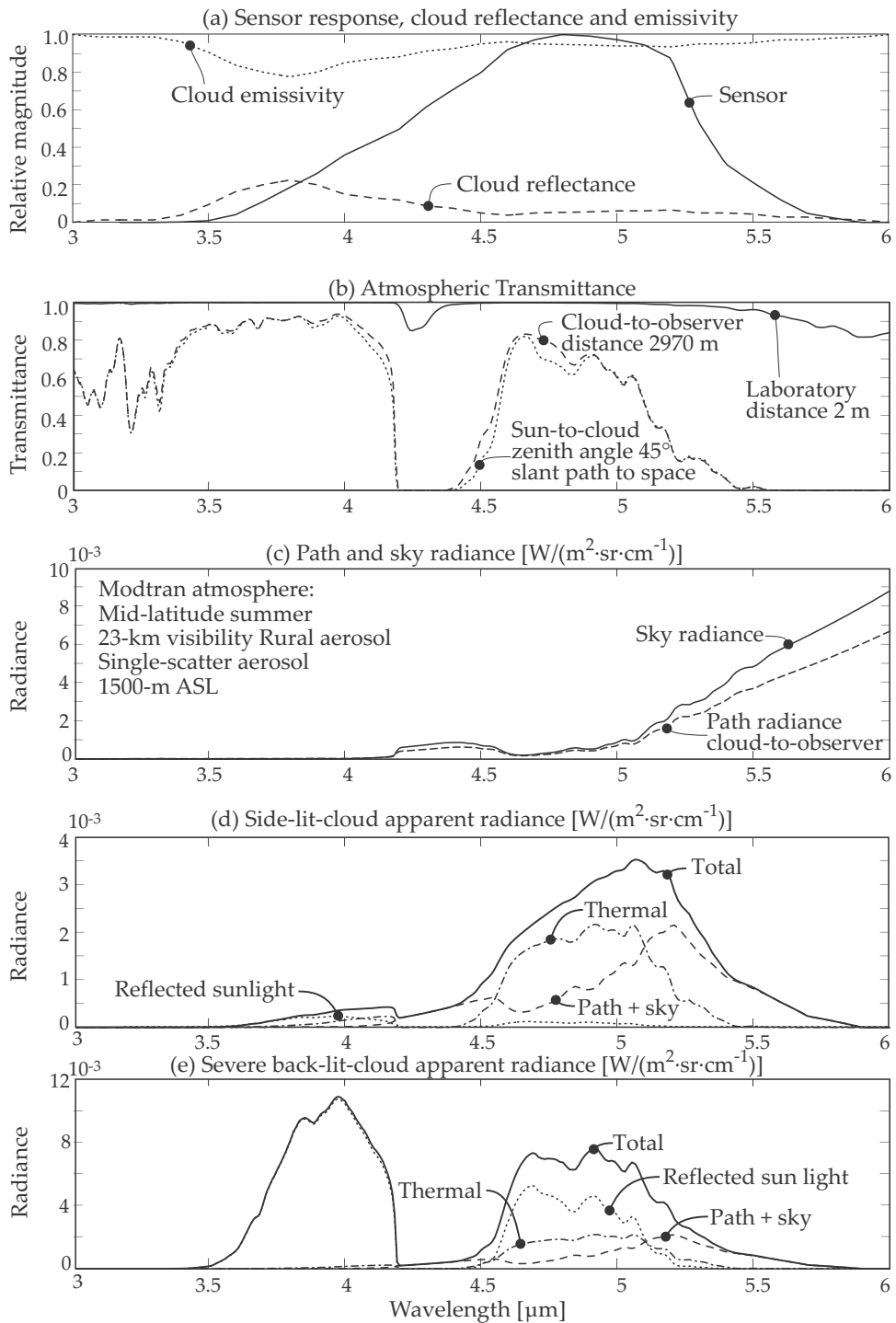


Figure 8.9 Cloud model analysis spectral input data and results.

Equation (8.14) was used to determine the silver-lining factor, using the measured cloud radiance values and cloud temperatures of 0 °C and 10 °C. The results are shown in Table 8.3. The results indicate that the model predicts side-lit cloud radiance 13% lower than the measurement. It is evident that the silver-lining factors are fairly robust against cloud temperature variations. These silver-lining factors compare favorably with silver-lining factors in the visual spectral band, which can be up to 70 for view angles close to the sun.

8.8.3 Relative contributions to the cloud signature

It is interesting to observe the relative contributions of the self emission and path radiance to the reflected sunlight. Figures 8.9(d) and 8.9(e) indicate each component individually. Path radiance plays a large role in the magnitude of the signature. For side-lit clouds the thermal component dominates, and for back-lit clouds the scattered or reflected sunlight component dominates.

8.9 Case Study: Contrast Inversion/Temperature Cross-Over

The temperature of objects in a natural scene depends on the diurnal heat flow, mainly following solar and seasonal patterns. Consider the thermal history of objects in an open terrain:

1. Early in the morning, all objects have cooled down by radiation, conduction, and convection. Temperature contrast is low.
2. At sunrise, the absorbed sunlight raises the surface temperatures of some objects very quickly. During the few minutes after sunrise, the infrared signatures in the scene change very quickly. Some materials have low thermal conductance between the surface layers and deeper thermal capacitance. A thin layer on the surface of these objects heats up very quickly because the heat flow into the surface is higher than the heat flow out of the surface (by radiation and conduction). The effect is quite dramatic during low-contrast conditions just before sunrise. This is a surface-temperature rise effect, not a bulk temperature effect.
3. By mid-morning, the objects have received a strong injection of energy from the sun. Some objects (e.g., rocks, metal, roads) respond quicker to the solar influx and heat up more rapidly than other objects (vegetation, wet soil). The result is that the average bulk temperature is still relatively low, but the thermal contrast is high.

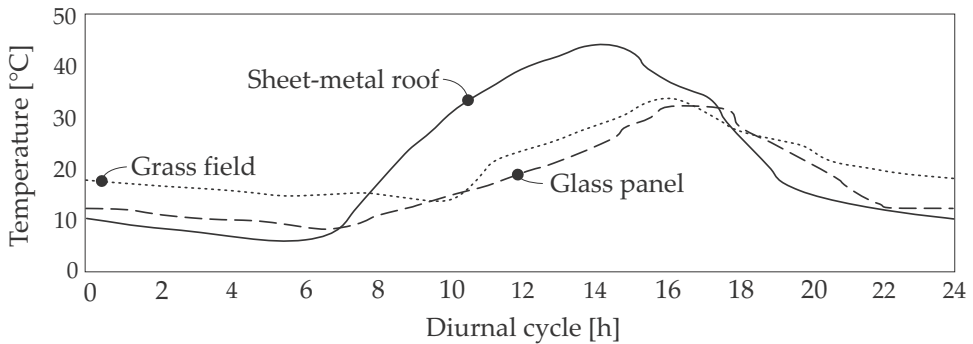


Figure 8.10 Temperature inversion caused by time lag in object temperatures.

4. In mid-afternoon, all objects are hot, having absorbed heat all morning. The contrast between the objects is, however, decreasing because most objects have been heated by now.
5. In early evening, the objects with low heat capacity have lost much of their energy by radiation, whereas objects with high heat capacity are still relatively warm. As the temperature decreases, the contrast increases slightly until all objects have lost most of their excess energy.

The temporal behavior of each object is unique because objects have different heat capacities, thermal conductances, and heat exchange with other objects. Each object therefore has a unique thermal history.

At least twice during this diurnal cycle, two objects may have the same radiance and hence have zero contrast with respect to each other — the edges between the objects disappear in a thermal image. This is known as thermal or temperature crossover. In practice, in complex or rich scenes, emissivity differences and small temperature differences between objects remain, rendering at least some part of the object visible. However, if the scene does not contain much detail (such as in a sand desert), even the distinction between the sky and ground may fade.

The morning crossover effect is much stronger than the evening crossover effect because by the early evening there are still significant amounts of heat stored in the bulk of the objects' bodies, leading to a wider temperature distribution in the scene.

8.10 Case Study: Thermally Transparent Paints

A paint with the required visual properties, but which is transparent in the thermal bands, could be useful in a multi-spectral signature management

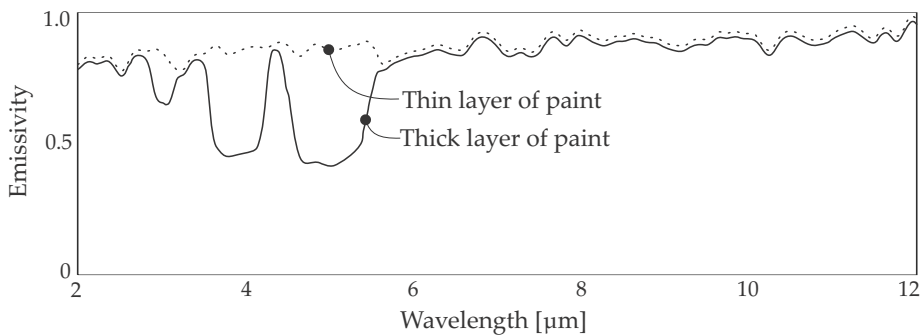


Figure 8.11 Emissivity of a metal plate painted with a thermally transparent paint.

(camouflage) design. The object would have the appropriate visual colors, but it would reflect the ambient surroundings' infrared signature from a shiny metal subsurface.

One such thermally transparent paint was characterized after having been applied to a clean, polished aluminium plate. The resulting emissivity value is shown in Figure 8.11. Note the low emissivity in the 3–5- μm spectral band arising from the fact that the paint is transparent in the 3–5- μm spectral band, and the polished subsurface is observed. Note also that the 8–12- μm spectral band emissivity is still too high to be of practical value.

The dependency of emissivity on paint thickness indicates that the paint base is not fully transparent. A thick coating ($> 150 \mu\text{m}$) will effectively cover any metal surface underneath but with emissivity equal to other paints.

8.11 Case Study: Sun-Glint

The 3–5- μm spectral band contains sunlight, as well as thermally emitted energy. Figure 8.12 shows a measurement of sun-glint in a harbor, in the direction of the sun, on a calm day. The probability of observing this severe sun-glint is relatively low because the sun elevation angle must be low the sensor must be pointed in the direction of the sun, and the sea must be calm. Most observations will result in less sun-glint than is demonstrated here. It is evident in Figure 8.12 that sunlight dominates at the shorter wavelengths. Sun-glint is also sometimes visible in the 8–12- μm spectral band but at reduced levels relative to the 3–5- μm spectral and visual bands.

Sun-glint is less common in nonmaritime scenarios but does occur from specularly reflecting objects, such as galvanized steel roof plates, inland bodies of water, ice caps, or similar shiny surfaces.

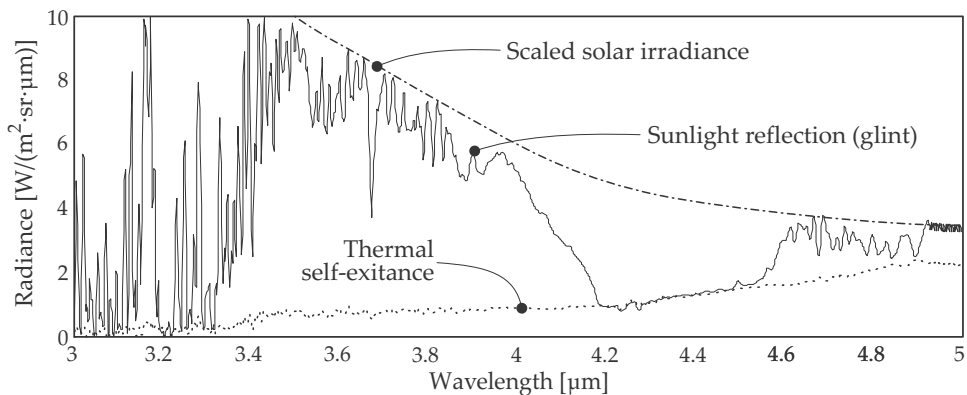


Figure 8.12 Sunlight reflection of the sea surface (glint), sun at low elevation, looking in the direction of the sun.

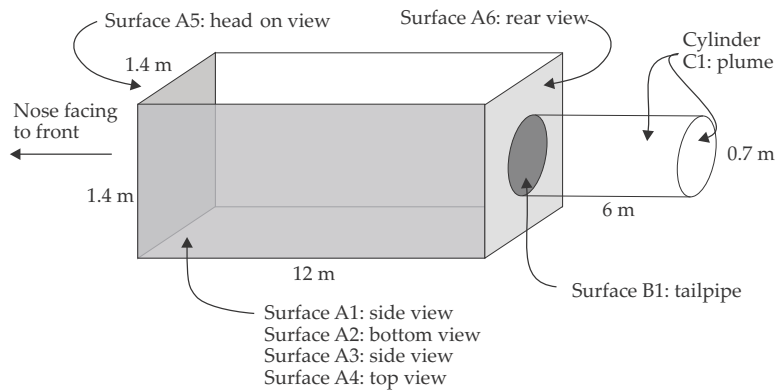
Bibliography

- [1] Pyradi team, "Pyradi Radiometry Python Toolkit," <http://code.google.com/p/pyradi>.
- [2] Willers, C. J., Willers, M. S., and Lapierre, F. D., "Signature modelling and radiometric rendering equations in infrared scene simulation systems," *Proc. SPIE* **8187**, 81870R (2011) [doi: 10.1117/12.903352].
- [3] Wolfe, W. L. and Zissis, G., *The Infrared Handbook*, Office of Naval Research, US Navy, Infrared Information and Analysis Center, Environmental Research Institute of Michigan (1978).
- [4] Accetta, J. S. and Shumaker, D. L., Eds., *The Infrared and Electro-Optical Systems Handbook (8 Volumes)*, ERIM and SPIE Press, Bellingham, WA (1993).
- [5] Palmer, J. M. and Grant, B. G., *The Art of Radiometry*, SPIE Press, Bellingham, WA (2009) [doi: 10.1117/3.798237].
- [6] Jacobs, P. A., *Thermal Infrared Characterization of Ground Targets and Backgrounds*, SPIE Press, Bellingham, WA (1996) [doi: 10.1117/3.651915].
- [7] Mahulikar, S. P., Sonawane, H. R., and Rao, G. A., "Infrared signature studies of aerospace vehicles," *Progress in Aerospace Sciences* **43**(7-8), 218–245 (October–November 2007).
- [8] Ferwerda, J. G., Jones, S. D., and Reston, M., "A free online reference library for hyperspectral reflectance signatures," SPIE News-Room (Dec 2006).

- [9] Johansson, M. and Dalenbring, M., "Calculation of IR signatures from airborne vehicles," *Proc. SPIE* **6228**, 622813 (2006) [doi: 10.1117/12.660108].
- [10] Hudson, R. D., *Infrared System Engineering*, Wiley-Interscience, New York (1969).
- [11] Roblin, A., Baudoux, P. E., and Chervet, P., "UV missile plume signatures model," *Proc. SPIE* **4718**, 344–355 (2002) [doi: 10.1117/12.478822].
- [12] Neele, F. and Schleijpen, R., "UV missile plume signatures," *Proc. SPIE* **4718** (2002).
- [13] Mahmoodi, A., Nabavi, A., and Fesharaki, M. N., "Infrared image synthesis of desert backgrounds based on semiempirical thermal models," *Optical Engineering* **40**(2), 227–236 (2001) [doi: 10.1117/1.1337037].
- [14] Rapanotti, J., Gilbert, B., Richer, G., and Stowe, R., "IR sensor design insight from missile plume prediction models," *Proc. SPIE* **4718** (2002) [doi: 10.1117/12.478816].
- [15] Magalhães, L. B. and Alves, F. D. P., "Estimation of radiant intensity and average emissivity of Magnesium/Teflon/Viton (MTV) flares," *Proc. SPIE* **7662**, 766218 (2010) [doi: 10.1117/12.850617].
- [16] van den Bergh, J. H. S., "Specular reflection," private communication (2004).
- [17] Michalski, L., Eckersdorf, K., Kucharski, J., and McGhee, J., *Temperature Measurement*, 2nd Ed., John Wiley and Sons, New York (2001).
- [18] Strojnik, M., Paez, G., and Granados, J. C., "Flame thermometry," *Proc. SPIE* **6307**, 63070L (2006) [doi: 10.1117/12.674938].
- [19] Liberman, M., *Introduction to Physics and Chemistry of Combustion*, Springer, Berlin (2008).
- [20] Glassman, I., *Combustion*, Academic Press, San Diego, CA (1987).

Problems

- 8.1 Consider a simple model for an aircraft consisting of three opaque components: a rectangular solid (the fuselage), a disc (the tailpipe), and a solid cylinder (the plume), as shown in the following figure:



with the following properties:

Surface	Temperature	Emissivity
A1 fuselage side	60	0.8
A2 fuselage bottom	60	0.8
A3 fuselage side	60	0.8
A4 fuselage top	60	0.8
A5 front view	75	0.8
A6	75	0.8
B1	600	0.8
C1	500	spectral
	°C	

Use Equation (D.4) to model the plume spectral emissivity with ($\tau_s = 0$, $\tau_p = 0.5$, $\lambda_c = 4.33 \mu\text{m}$, $\Delta\lambda = 0.45 \mu\text{m}$, and $s = 6$).

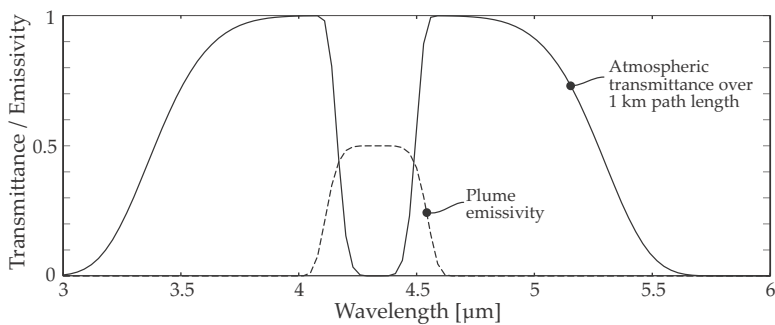
The atmospheric transmittance at a range of 1000 m is given by $\tau_{1000} = \tau_{1\lambda} (1 - \tau_{2\lambda})$, where the values $\tau_{1\lambda}$ and $\tau_{2\lambda}$ are calculated using Equation (D.4). The $\tau_{1\lambda}$ parameters are: ($\tau_s = 0$, $\tau_p = 1$, $\lambda_c = 4.33 \mu\text{m}$, $\Delta\lambda = 2 \mu\text{m}$, and $s = 6$). The $\tau_{2\lambda}$ parameters are: ($\tau_s = 0$, $\tau_p = 1$, $\lambda_c = 4.33 \mu\text{m}$, $\Delta\lambda = 0.35 \mu\text{m}$, and $s = 6$).

Aspect angle is measured in the horizontal plane from the nose; a zero aspect angle is looking at the aircraft nose; an aspect angle of π is looking at the aircraft tail.

Do all spectral calculations in the spectral range from $3 \mu\text{m}$ to $6 \mu\text{m}$ with an increment of $0.05 \mu\text{m}$.

You can ignore reflected sunlight and the blue sky.

All solid surfaces on the plume have the same radiometric properties.



8.1.1 Calculate the transmittance from the above equation and confirm that it agrees with the graph. Calculate and plot the attenuation coefficient. Then calculate and plot the transmittance at the following ranges: 500 m, 1000 m, 2000 m, and 5000 m. [3]

8.1.2 Draw plots for each component (surface or plume), showing the irradiance with the atmosphere present for all of the ranges on the same plot. [3]

Calculate the total (integrated) irradiance received by a sensor (from all of the surfaces facing the sensor) at all above ranges from the aircraft, for aspect angles $\theta \in \{0, \pi/2, \pi\}$ rad, with no atmosphere present. If more than one surface is visible, calculate and show the contributions separately, and also the sum of all contributions. [3]

8.1.3 Draw plots for each component (surface or plume), showing the irradiance with the atmosphere not present, for all of the ranges on the same plot. [3]

Calculate the total (integrated) irradiance received by a sensor (from all of the surfaces facing the sensor) at all above ranges from the aircraft, for aspect angles $\theta \in \{0, \pi/2, \pi\}$ rad, with the atmosphere present. If more than one surface is visible, calculate and show the contributions separately, and also the sum of all contributions. [3]

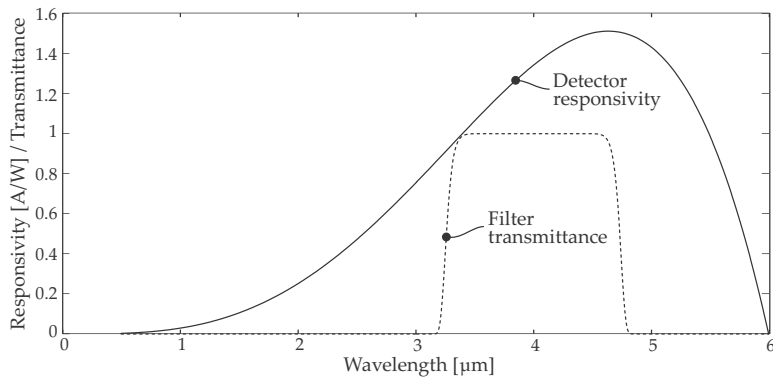
8.2 An InSb-based sensor is pointed toward a spherical cloud in the sky at a range of 1000 m. The cloud is illuminated by the sun, from behind the sensor. Ignore any obscuration of the sunlight by the sensor or the earth.

The infinitely-large cloud has a temperature of 10°C , and an emissivity of 0.8.

The sensor has a FOV of 10^{-5} sr. The sensor FOV is completely filled by the cloud. The sensor optical aperture diameter is 50 mm.

The sensor detector responsivity can be modeled by Equation (D.5) with ($\lambda_c = 6 \mu\text{m}$, $k = 20$, $a = 3.5$, and $n = 4.3$).

Use Equation (D.4) to model the sensor filter transmittance with ($\tau_s = 0$, $\tau_p = 1$, $\lambda_c = 4 \mu\text{m}$, $\Delta\lambda = 1.5 \mu\text{m}$, and $s = 20$).



You may assume unity atmospheric transmittance and zero path radiance.

The task is to calculate the total signature of the cloud as observed by the sensor in the MWIR spectral band. The signature comprises the reflected sunlight as well as the thermal self-emittance.

- 8.2.1 Write a mathematical formulation describing the detector current; include flux transfer, detector response, etc. Describe all elements in the model and provide the relevant numerical values for all parameters. [3]
- 8.2.2 Apply the Golden Rules to the mathematical formulation. [3]
- 8.2.3 Build a computer model of the problem. Describe the structure of the model and provide all numeric values (spectral and scalar). Use a wavelength increment of $0.01 \mu\text{m}$ in the spectral range 0.3 to $6 \mu\text{m}$. [4]
- 8.2.4 Compile graphs of the spectral irradiance from the sources on the entrance aperture of the sensor. The graphs must show the two sources separately, as well as the sum of the two, with no detector or filter weighting (i.e., in front of the sensor). Compile three sets of graphs: (1) unfiltered irradiance (no detector or filter), (2) detector-weighted irradiance (no filter), and (3) detector and filter weighted. Use the normalized detector response for spectral weighting. [6]
- 8.2.5 Calculate the current through the detector when it is viewing the cloud if no optical filter is present. [2]

- 8.2.6 Calculate the current through the detector when it is viewing the cloud with the optical filter present. [2]
- 8.2.7 Comment on the use of the 3–6- μm spectral band for observations during day- and nighttime. What consideration should be made when performing temperature measurement in this spectral band? [2]
- 8.3 A thin circular disk has a temperature of 300 K on the one side and a temperature of 1000 K on the other side. The emissivity on both sides is the same with a value of 1. The disk diameter is 1 m. Assume an atmospheric transmittance of unity. The disk is viewed against an infinitely large background with a uniform spatial temperature distribution. Consider two scenarios: a background temperature of 270 K and a background temperature of 330 K.
- A sensor is located at a range of 1000 m from the disk. The sensor spectral response is unity in the 3–5- μm spectral band and zero elsewhere. The sensor FOV is 2 mrad full-apex angle.
- The objective with this investigation is to plot the target contrast intensity from all view directions (similar to the graphs shown in Figure 8.1) for both background temperatures.
- Prior to calculation, try to visualize the spherical intensity, and draw a spherical contrast intensity diagram in freehand.
- Calculate and plot the polar contrast intensity of the disk against the two backgrounds, in the xy , yz , and zx planes. [6]
- Use the tools in `pyradi`¹ (or any other tool of your choice) to calculate and plot the three-dimensional spherical contrast intensity of the disk against the two backgrounds. [10]
- 8.4 Extend the one-dimensional wedge light trap described in Problem 3.16.2 to two dimensions, with the wedge lines running along the x axis. The light trap has a temperature of 1000 K. Calculate and plot the hemispherical radiance distribution of the light trap (use `pyradi`¹ or any other tool). [10]

Chapter 9

Electro-Optical System Analysis

*It doesn't matter how beautiful your theory is,
it doesn't matter how smart you are.
If it doesn't agree with experiment, it's wrong.*
Richard P. Feynman

Introduction

In this chapter different electro-optical systems are briefly defined and analyzed to demonstrate the application of the radiometric modeling techniques developed in the early chapters of the book. In the work presented here, the emphasis is on the radiometry and methodology, rather than on the detail parameters. Some of the case studies may appear somewhat contrived, but these are still useful hands-on training material. In any real design, considerably more effort will be expended: the level of detail will be deeper, and the models much more comprehensive. One example of a more comprehensive application is the simulation system described in Appendix B.

9.1 Case Study: Flame Sensor

The flame sensor must detect the presence or absence of a flame in its FOV. The sensor is pointed to an area just outside a furnace vent, against a clear-sky background. The sensor must detect a change in signal indicating the presence of a flame at the vent. This problem was defined to illustrate the calculation of spectral integrals. This case study is also a worked example in Section D.5.1 and on the pyradi toolkit website.¹

The sensor has an aperture area of $7.8 \times 10^{-3} \text{ m}^2$ and a FOV of 1×10^{-4} sr. The sensor filter spectral transmittance is shown in Figure 9.1(a). The spectral transmittance can be calculated with Equation (D.4) and parameters ($\tau_s = 0.0001$, $\tau_p = 0.9$, $s = 12$, $\Delta\lambda = 0.8 \text{ } \mu\text{m}$, $\lambda_c = 4.3 \text{ } \mu\text{m}$). The InSb

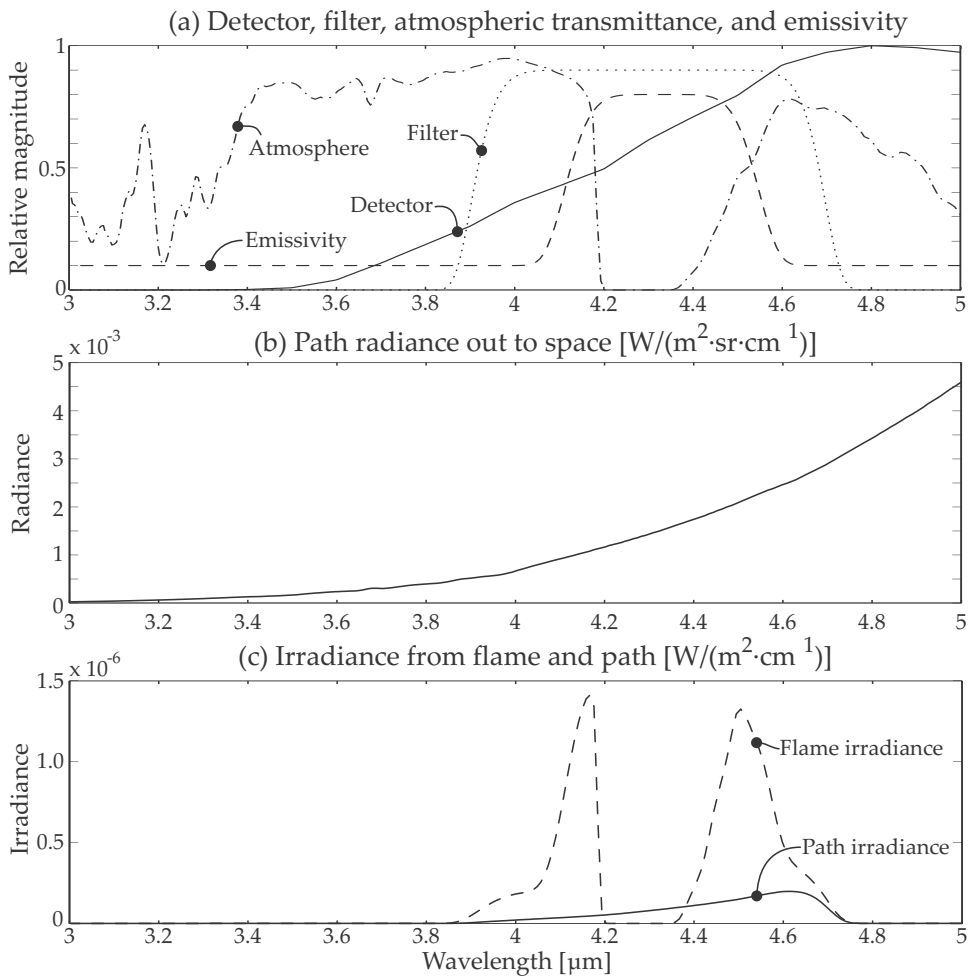


Figure 9.1 Flame sensor spectral variables and results.

detector has a peak responsivity of 2.5 A/W and spectral response shown in Figure 9.1(a). The preamplifier transimpedance is 1×10^4 V/A.

The flame area is 1 m². The flame temperature is 1000 °C. The emissivity is 0.1 over most of the spectral band due to carbon particles in the flame. At 4.3 μm there is a strong emissivity rise due to the hot CO₂ in the flame; see Figure 9.1(a). The emissivity can be calculated with Equation (D.4) and parameters ($\tau_s = 0.1$, $\tau_p = 0.7$, $s = 6$, $\Delta\lambda = 0.45$ μm , $\lambda_c = 4.33$ μm).

The distance between the flame and the sensor is 1000 m. The atmosphere is similar to the MODTRANTM Tropical climatic model. The path is oriented such that the sensor stares out to space at a zenith angle of 88.8 deg. The components' spectral transmittance is shown in Figure 9.1(a) and the path radiance in Figure 9.1(b).

Table 9.1 Summary results for flame sensor analysis.

Characteristic	Value	Unit
Path irradiance	0.054	mW/m ²
Path voltage	10.4	mV
Flame irradiance	0.329	mW/m ²
Flame voltage	64	mV
Flame + path irradiance	0.382	mW/m ²
Flame + path voltage	75	mV

The spectral peak in the flame emissivity and the dip in atmospheric transmittance are both centered around the 4.3- μm CO₂ band. The strong spectral variations in both target signature and atmospheric transmittance necessitates multi-spectral calculation (see Section 2.6.5).

From Equation (6.16), the signal caused by the atmospheric path radiance is given by

$$v_{\text{path}} = k \hat{\mathcal{R}} Z_t \Omega A_1 \int_0^\infty L_{\text{path}\lambda} \mathcal{S}_\lambda d\lambda, \quad (9.1)$$

where $\mathcal{S} = \widetilde{\mathcal{R}}_\lambda \tau_{f\lambda}$. Note that the path radiance terms do not have an atmospheric transmittance factor because the radiance is the net effect of the atmosphere [see Equation (4.10)]. The signal caused by the flame is given by

$$v_{\text{flame}} = \frac{k \hat{\mathcal{R}} Z_t A_0 A_1}{R_{01}^2} \int_0^\infty \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda, \quad (9.2)$$

where the variables are defined as for Equation (6.16). A_0 is the flame area because $\cos \theta_0 = 1$, and A_1 is the sensor aperture area. These equations were evaluated as described in Section D.5.1, and yielded the results shown in Table 9.1.

It is clear that the flame signal is several times larger than the path radiance signal, even though the flame only fills 1% of the sensor FOV. The severity by which the atmosphere attenuates the CO₂ exitance from the flame is shown in Figure 9.1(c).

9.2 Case Study: Object Appearance in an Image

Section 7.6 developed the theory for observing a simple target against a background through a radiant medium. This section describes the model used to calculate practical values for this scenario. The section closes with a

prediction of the meteorological range of the medium given the irradiance contrast versus range of white and black targets.

An opaque target of fixed size is viewed against an infinitely large opaque background at path lengths from short range to very long range. The target and background are parallel flat surfaces located at the same distance from the sensor and collectively fill the complete FOV. The target and background surfaces are both facing the sensor, with near-horizontal normal vector. The sun is located at zero zenith angle (i.e., vertically above the target).

A variation of Equation (8.1) is used, ignoring transmitted background flux and diffusely reflected sky and background radiance. The retained terms are thermal self-exittance, reflected sunlight, and atmospheric path radiance. The target and background BRDF, $f_r(\theta_s, \theta_v)$ are assumed Lambertian. For this analysis the pixel FOV observes a combination of target radiance and background radiance, with the ratio between target and background varying with range:

$$\begin{aligned}
 E_S = & \overbrace{Y\omega_p \int_0^\infty \epsilon_{t\lambda}(\theta_v) L_\lambda(T_t) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{target thermally emitted } L_{\text{self T}}} \\
 & \overbrace{(1-Y)\omega_p \int_0^\infty \epsilon_{b\lambda}(\theta_v) L_\lambda(T_b) \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{background thermally emitted } L_{\text{self B}}} \\
 & \overbrace{+ \psi \cos \theta_s Y\omega_p \int_0^\infty (1 - \epsilon_{t\lambda}) \epsilon_{s\lambda} L_\lambda(T_s) \tau_{s0\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{target reflected sun } L_{\text{ref sun T}}} \\
 & \overbrace{+ \psi \cos \theta_s (1-Y)\omega_p \int_0^\infty (1 - \epsilon_{b\lambda}) \epsilon_{s\lambda} L_\lambda(T_s) \tau_{s0\lambda} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda}^{\text{background reflected sun } L_{\text{ref sun B}}} \\
 & \overbrace{+ \omega_p \int_0^\infty L_{\text{path}\lambda} \mathcal{S}_\lambda d\lambda}^{\text{atmospheric path radiance } L_{\text{path}}} \quad , \quad (9.3)
 \end{aligned}$$

where E_S is the irradiance at the sensor entrance aperture in the pixel FOV ω_p , $\epsilon_{t\lambda}$ is the target emissivity, T_t is the target temperature, $\epsilon_{b\lambda}$ is the background emissivity, and T_b is the background temperature. If the target area is A_t and the distance between the sensor and the target is R , then $Y = A_t / (R^2 \omega_p)$, and $0 \leq Y \leq 1$ is the fraction of the FOV that is filled with the target. The remaining terms are defined in Table 8.1.

The first analysis using Equation (9.3) is to investigate the relative contributions of the target, background radiance, and path irradiance to

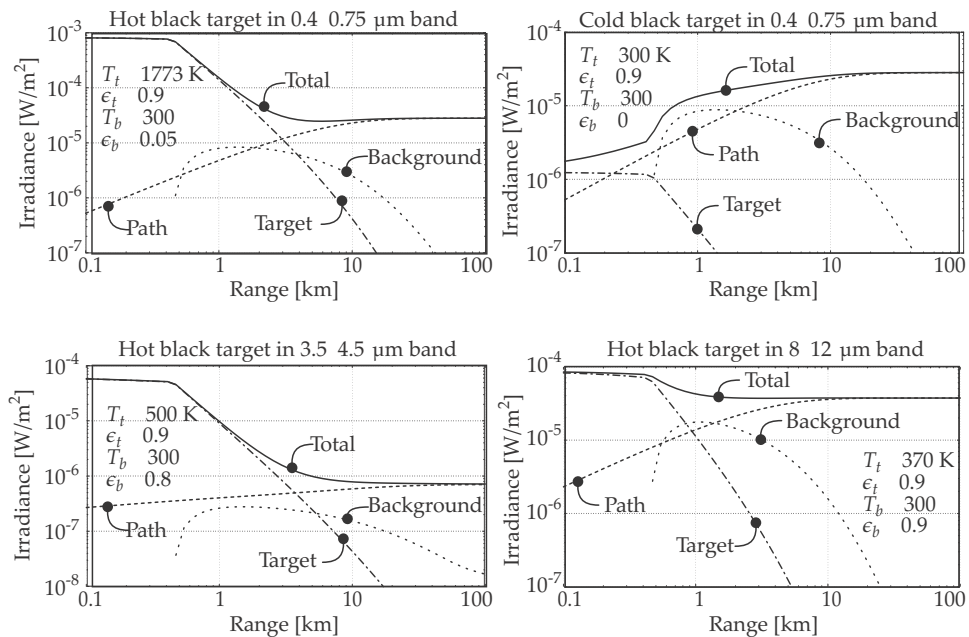


Figure 9.2 Pixel irradiance as function of object-to-observer distance for varying target temperatures and sensor spectral bands.

the total irradiance in the pixel. The analysis was done for the following MODTRAN™ atmosphere: Tropical profiles, 23-km visibility Rural aerosol ('MIE Generated' aerosol phase function), executed in the 'Radiance with Scattering' mode, with multiple-scattering for flux at the observer. The Isaac's two-stream multiple-scattering algorithm is used. The observer is located at sea level, viewing a slant path with 88-deg zenith angle (near horizontal). The pixel FOV is 1 μsr, and the target area is 0.2 m². The target and background temperature and emissivity properties are indicated on the graphs in Figure 9.2. The code to calculate these graphs is included in Section D.5.3.

From Figure 9.2 it is evident that the target determines the pixel irradiance for targets at close range, whereas the atmospheric path radiance dominates the signature for targets at long range. The 'cold-black' target in the visual band presents an interesting case in that at close target range the pixel irradiance is low, because there is no sunlight reflection from the target. At intermediate ranges the white reflective background dominates the signature. At the risk of generalizing based on specifics, at least for the test cases shown here, the atmospheric path radiance dominates from 10 km onwards. In the visual and LWIR spectral bands the path radiance magnitude increase exceeds an order of magnitude over the ranges considered. In the MWIR spectral band, this increase is considerably less than

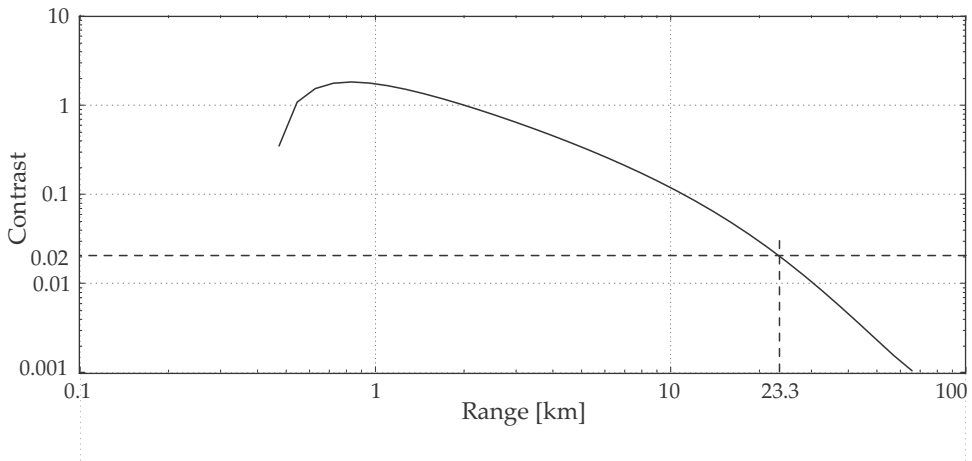


Figure 9.3 Contrast between target pixel and adjacent background pixel in the visual spectral band.

an order of magnitude, indicating a more-favorable scenario.

In the second analysis, the meteorological (Koschmieder) range is calculated from the pixel irradiance values. Now the background around the target becomes the object of interest because the small area target has an insignificant contribution beyond 10 km. The scenario shown in the top right graph in Figure 9.2 is a black target against a *white* background. A similar graph was also calculated for the target against a *black* background. Using Equation (4.14) and the two (black and white) background irradiance graphs, the contrast shown in Figure 9.3 was calculated using

$$C_R = \frac{E_{\text{white background}} - E_{\text{black background}}}{E_{\text{black background}}} . \quad (9.4)$$

Equation (9.4) indicates that the predicted meteorological range depends on the magnitude of the incident solar irradiance $E_{\text{white background}}$. Using the Koschmieder threshold contrast definition in Section 4.6.10, the meteorological range was calculated as 23.3 km but for a very specific background orientation. If the background surface is horizontal, fully facing the sun, $\cos \theta_s = 1$, and $E_{\text{white background}}$ is large, leading to a very long meteorological range. For a background surface tilted such that angle between the surface normal vector and sun direction is 84 deg (near-vertical surface), $\cos \theta_s = 0.1$, and the desired result is obtained. Is this model manipulation reasonable? Consider the aviation definition of meteorological range: the path should be near the horizontal, observing objects on the horizon. Under this condition, the visible object surfaces will be closer to the vertical than to the horizontal. Hence, the model should make provision for near-vertical surfaces and then yield reasonable results.

Table 9.2 Solar-cell measurements.

	Sun	Lamp
Open-circuit voltage [V]	1.59	1.2
Short-circuit current [A]	0.107	0.06
Load line for fan		
Load voltage [V]	0.9	0.54
Load current [A]	0.09	0.054

9.3 Case Study: Solar Cell Analysis

9.3.1 Observations

In a simple experiment, an inexpensive, toy silicon solar-cell panel was used to power a small fan when illuminated by a lamp and the sun. The panel had eight elements connected in an unknown network. The solar cell had a fill factor of about 50%, estimated by inspection. The estimated solar panel spectral responsivity is shown in Figure 9.4. The silicon solar cells did not have an antireflection coating.

The MODTRANTM Tropical atmosphere, with 30-deg zenith angle slant path to space and 20-km visibility Urban aerosol attenuation, was used in the analysis. The spectral transmittance is shown in Figure 9.4.

The panel was illuminated by an incandescent lamp and the sun. In each case three load conditions were recorded: (1) the open-circuit voltage (i.e., no load), (2) the short-circuit current (i.e., no voltage), and (3) the fan load. For the two measurements, the atmospheric transmittance was assumed spectrally constant at $\tau_{\text{lamp}} = 1$ and $\tau_{\text{sun}} = 0.7$. The recorded values are shown in Table 9.2.

The panel was illuminated by a 60-W incandescent lamp (color temperature of approximately 2650 K) at a perpendicular distance of 60 mm from the center of the panel (Figure 9.4). The lamp filament emissivity in the visual spectrum can be assumed to be 0.5. The lamp was housed in a black painted fitting with a reasonable (but not perfect) matte finish. There was a slight obscuration of the panel corners by the lamp housing.

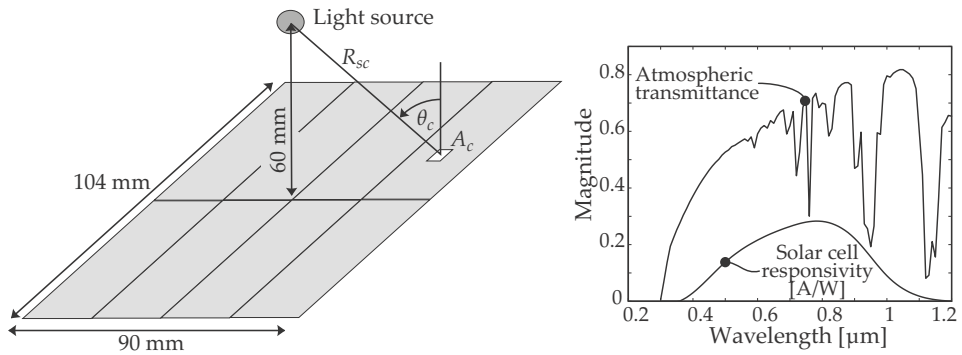


Figure 9.4 Solar panel and lamp geometry, spectral atmospheric transmittance, and spectral solar cell responsivity.

9.3.2 Analysis

9.3.2.1 Solid angles and source areas

The sun area is calculated by simple mathematics: $A_{\text{sun}} = \pi R_{\text{sun}}^2 = 1.5 \times 10^{18} \text{ m}^2$. The solid angle of the solar panel as seen from the sun is approximated by the area of the panel divided by the distance to the sun squared because the panel area is small compared to the distance to the sun:

$$\Omega_{\text{panel from sun}} = \frac{104 \times 90 \times 10^{-6}}{(149 \times 10^9)^2} = 421 \times 10^{-27} \text{ sr.} \quad (9.5)$$

The lamp is modeled as an isotropic point target with no $\cos \theta$ weighting. The solid angle of the solar panel, as seen from the lamp, is calculated using the integral in Equation (2.13):

$$\begin{aligned} \Omega_{\text{panel from lamp}} &= \int_W \int_D \frac{dw \, dd}{H^2} \left(\frac{H}{\sqrt{w^2 + d^2 + H^2}} \right)^3 \\ &= 1.63 \text{ sr.} \end{aligned} \quad (9.6)$$

The integral is determined over the size of the solar panel: $-0.045 \text{ m} \leq h \leq 0.045 \text{ m}$, and $-0.052 \text{ m} \leq h \leq 0.052 \text{ m}$. For more details on how to calculate the integral, see Section D.5.8.

The lamp area is calculated with a little trick: the Stefan–Boltzmann law (Section 3.1.3) states that the total exitance by a blackbody is given by $M(T) = 5.67 \times 10^8 T^4$, but the total power is known to be 60 W (the lamp rating). For this calculation, assume that the only energy loss is through radiation with no heat loss due to convection or conduction through the filaments' stem wires. From $\Phi = M(T)A_{\text{lamp}}$, it follows that the total

radiating surface area must be

$$A_{\text{lamp}} = \frac{60}{5.67 \times 10^8 (2650)^4} = 21 \times 10^{-6} \text{ m}^2, \quad (9.7)$$

corresponding to a square area of about 4.6 mm by 4.6 mm, which seems reasonable considering the length and coiled shape of the filament. How robust is the lamp temperature assumption and its effect on the solution? The exitance varies with temperature to the fourth power, so a variation in temperature could have a significant effect in the result. If the lamp temperature is 2850 K, the filament area drops only a little to 4 mm by 4 mm. Thus, the result is reasonably robust against error in lamp-temperature estimation.

9.3.2.2 Radiometry calculations

Following the workflow in Section 6.6 the solar cell current can be written as follows:

$$\begin{aligned} i_{ph} &= f_{\text{fill}} \hat{\mathcal{R}} \int_{\lambda=0}^{\infty} \int_{\text{source}} \int_{\text{cell}} \frac{\epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_{\lambda} d\lambda dA_0 dA_1 \cos \theta_0 \cos \theta_1}{R_{01}^2} \\ &= f_{\text{fill}} \hat{\mathcal{R}} \int_{\lambda=0}^{\infty} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_{\lambda} d\lambda \int_{\text{source}} dA_0 \int_{\text{cell}} \frac{dA_1 \cos \theta_0 \cos \theta_1}{R_{01}^2} \quad (9.8) \\ &= f_{\text{fill}} I_{\text{eff}} A_0 \Omega_1, \end{aligned}$$

where i_{ph} is the photon-induced current, f_{fill} is the fill factor (how much of the panel is active detector area), $\hat{\mathcal{R}}$ is the detector responsivity scaling factor with units [A/W], the spatial integrals and cosine terms are important in the case of the lamp, A_0 is the area of the source (sun or lamp), A_1 is the total physical area of the solar panel, and $\mathcal{S} = \widetilde{\mathcal{R}}_{\lambda} \tau_{f\lambda}$ is the solar-cell spectral responsivity.

The fill factor is the fraction of the panel that is able to convert optical power to electrical power. In this case, the panel consisted of a number of irregular pieces of silicon that filled only 50% of the area. This implies that only half of the physical area can respond to the flux and create electricity.

The lamp illumination did not cover the full area of the source because it was somewhat shielded by the lamp cover. It is estimated that 80% of the solar cell was illuminated.

Note that the solid angle calculated above is for the *whole* panel. However, the panel consists of eight cells, of which some are in series. It is shown below that the panel was wired as two parallel circuits, each of four cells in series. It follows that, for the purpose of current generation,

Table 9.3 Summary results for solar cell analysis.

Factor	Lamp	Sun	Unit
Area illuminated	80%	100%	
f_{fill}	0.5	0.5	
I_{eff}	1.84×10^4	1.64×10^6	A/(m ² ·sr)
A_0	2.15×10^{-5}	1.52×10^{18}	m ²
Ω_c	1.62/4	$4.22 \times 10^{-25}/4$	sr
Calculated current	0.064	0.131	A
Measured current	0.06	0.107	A
Difference	6.7%	22.4%	

the effective area, and therefore the solid angle, is 0.25 of the totals calculated above. The exact factor depends on where each cell is located in the panel, but for this calculation the value of 0.25 is used.

Using Equation (9.8), the calculated values shown in Table 9.3 are somewhat higher than the measured values. The difference could be attributed to errors in spectral response, fill factor, lamp temperature/area, or atmospheric transmittance.

The solar cell efficiency is the ratio of flux converted to electricity to the total incident flux:

$$\eta = \frac{\int_{\lambda=0}^{\infty} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} \mathcal{S}_{\lambda} d\lambda}{\int_{\lambda=0}^{\infty} \epsilon_{0\lambda} L_{0\lambda} \tau_{a\lambda} d\lambda}. \quad (9.9)$$

Using Equation (9.9), the efficiency was calculated. In this analysis the effect of atmospheric transmittance was ignored by setting $\tau_{a\lambda} = 1$. The value was found to be 1.4% for the lamp and 3.8% for the sun. It is evident that a very small portion of the incident flux is converted to electrical energy. The Shockley–Queisser limit for single-layer silicon p-n junction solar cells sets the theoretical limit at 33.7%. Current commercial and research solar cells achieve between 22% and 25% efficiency.

9.3.2.3 Configuration

Under ideal conditions, the highest output voltage is less than 0.58 V per single cell. Because the highest observed output voltage for this panel was 1.59 V, it is inferred that there must be four cells in series. In order to obtain the best load and generation distribution, the cells are connected as shown in Figure 9.5. Each cell acts as a current generator on its own. Small differences in photocurrents lead to small differences in cell voltages,

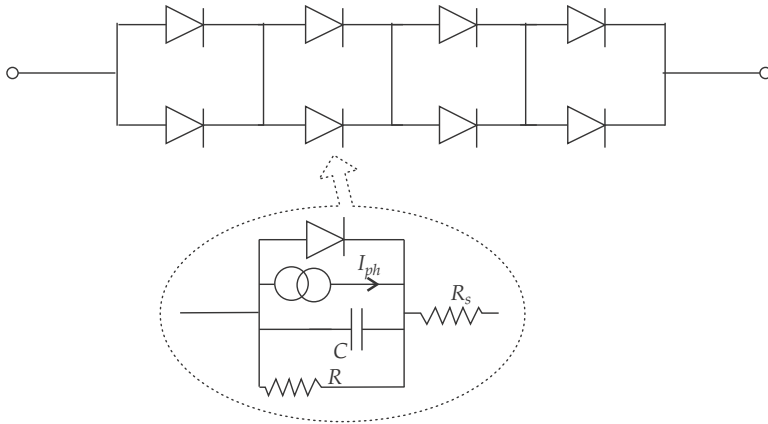


Figure 9.5 Solar panel circuit configuration.

which are equalized by voltage drop across the series resistors and cells.

9.3.2.4 Cell model

It is shown in Section 5.9.2 that the current generated by the solar cell is related to the cell voltage by the I-V equation of the form [Equation (5.128)]:

$$I_{\text{load}} = I_{\text{sat}} \left(e^{qV/(kT\beta)} - 1 \right) - I_{\text{ph}}. \quad (9.10)$$

Under short-circuit conditions it follows that

$$I_{\text{sc}} = I_{\text{sat}} (e^0 - 1) - I_{\text{ph}}, \quad (9.11)$$

and under open-circuit conditions

$$0 = I_{\text{sat}} \left(e^{qV_{\text{oc}}/(kT\beta)} - 1 \right) - I_{\text{ph}}. \quad (9.12)$$

Combining Equations (9.11) and (9.12),

$$0 = I_{\text{sat}} \left(e^{qV_{\text{oc}}/(kT\beta)} - 1 \right) + I_{\text{sc}}, \quad (9.13)$$

and combining Equations (9.11) and (9.10),

$$I_{\text{load}} = I_{\text{sat}} \left(e^{qV_{\text{load}}/(kT\beta)} - 1 \right) + I_{\text{sc}}. \quad (9.14)$$

Then, by combining Equations (9.13) and (9.14),

$$-I_{\text{load}} = I_{\text{sat}} \left(e^{qV_{\text{oc}}/(kT\beta)} - e^{qV_{\text{load}}/(kT\beta)} \right). \quad (9.15)$$

For $\beta = 1$ (ideal diffusion diode) it follows that $q/(kT\beta) = 38.65$, so that (from the sun short-circuit measurement) from Equation (9.13),

$$I_{\text{sat}} = \frac{0.054}{(e^{38.6 \times 0.397} - 1)} = 11 \times 10^{-9}. \quad (9.16)$$

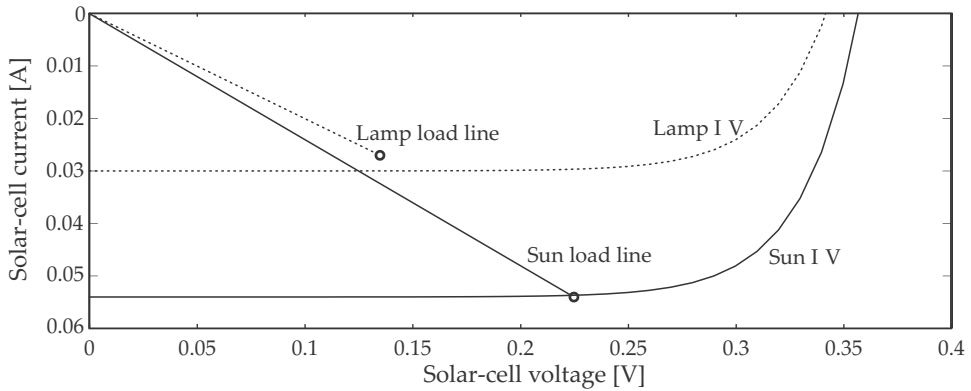


Figure 9.6 Solar cell I-V curve and load line for $I_{sc} = 0.03$ and $I_{sc} = 0.054$, lamp illumination, and sun illumination.

From the measurements and above calculations, the load current for the sun-illuminated cell is related to the cell voltage as

$$I_{\text{load}} = 11 \times 10^{-9} \left(e^{38.6 \times V_{\text{load}}} - 1 \right) - 0.054. \quad (9.17)$$

Using Equation (9.17), the sun-induced load current predicted for a cell voltage of 0.225 V is 0.045 A (per cell); this result is within 17% from the measured value of 0.9 V and 0.09 A (for the whole panel).

By similar analysis, the load current for the lamp illuminated cell is related to the cell voltage as

$$I_{\text{load}} = 276 \times 10^{-9} \left(e^{38.6 \times V_{\text{load}}} - 1 \right) - 0.03. \quad (9.18)$$

Using Equation (9.18), the lamp-induced load current predicted for a cell voltage of 0.135 V is 0.03 A (per cell). This result is within 7% from the measured value of 1.2 V and 0.06 A (for the whole panel). The two calculated values for I_{sat} were averaged by $\sqrt{I_{\text{sat sun}} * I_{\text{sat lamp}}} = 55 \times 10^{-9}$ A. Then, for any illumination condition with a short-circuit current of I_{sc} , the solar cell's load line is given by

$$I_{\text{load}} = 55 \times 10^{-9} \left(e^{38.6 \times V_{\text{load}}} - 1 \right) - I_{sc}. \quad (9.19)$$

Figure 9.6 shows the I-V curves for the two measurements described in the problem statement (for a single cell). The solar cell is illuminated by a lamp and the sun, while the same load is applied in both cases.

The analysis in this section indicates how a little information can be used to derive a model. Once the model is developed, the remaining challenge is validation that the model is indeed correct (Section 7.9). To some

extent this can be done with the information used to develop the model, but better confidence requires additional information and analysis.

9.4 Case Study: Laser Rangefinder Range Equation

In this section the range equation for a laser rangefinder is derived. Because the radiometry techniques developed in this book do not cover coherent sources, an explanation is in order. In this rangefinder application the laser is used as a source with very high radiance. Laser rangefinders operate on the principle that light travels approximately 300 mm distance in one nanosecond. The time elapsed between the transmission of the pulse and the reception of the pulse is used to determine the distance. This specific analysis does not require the laser to be considered as a coherent source, and hence the radiometry techniques in this book can be used here. These techniques cannot be readily used for problems concerned with spatial or temporal coherence.

In most laser rangefinders the transmitter and receiver are co-located and co-axial on the same optical path. If a laser pulse is directed to an object and reflected back from the object, the elapsed time between the departure and arrival of the reflected light pulse is an indication of the distance to the object. The objective with this analysis is to derive an expression for the SNR for the laser rangefinder. The SNR can then be used to investigate the effect of several design parameters on system performance. For another approach to the range equation for laser rangefinders, see Kaminsky.²

9.4.1 Noise equivalent irradiance

The noise equivalent irradiance (Section 6.7) in the receiver is given by

$$E_n = \frac{\sqrt{\Delta f A_d}}{D^* A_1 \tau_a}, \quad (9.20)$$

where A_d is the detector area, Δf is the noise bandwidth in the receiver, A_1 is the receiver aperture area, and τ_a is the receiver filter transmittance. The D^* values can include all of the relevant noise terms such as detector noise, amplifier noise, background induced noise, and system noise. The method whereby these noises are all combined into a single D^* is described in Section 5.3.12. For the purposes of this chapter, we will only work with a single D^* , assuming that all noise sources are incorporated in this value.

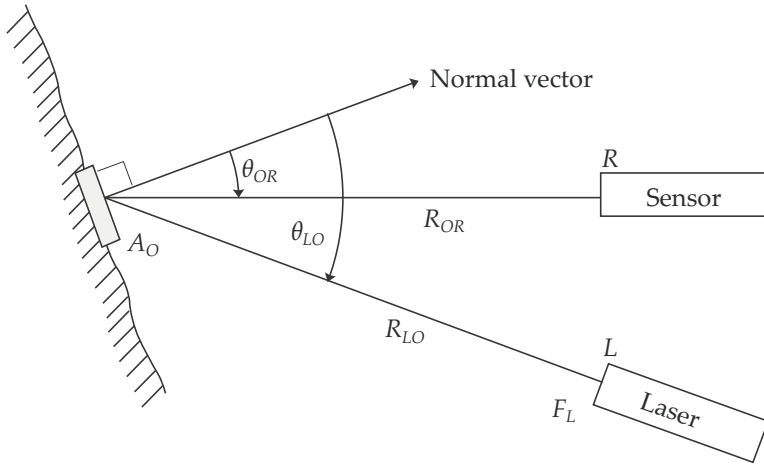


Figure 9.7 Laser rangefinder layout.

9.4.2 Signal irradiance

The geometrical relationship between the laser transmitter, the object, and the laser receiver is shown in Figure 9.7. The laser power or flux is denoted by Φ_L (in watts), the distance from the laser to the object is R_{LO} , and the distance from the object to the receiver is R_{OR} . The illuminated object-surface normal vector makes an angle θ_{LO} with the laser illumination direction, and an angle θ_{OR} with the receiver sightline direction.

Several simplifying assumptions are made in order to simplify the problem and emphasize the methodology. The receiver and transmitter fields are co-axial, and the receiver and transmitter are located at the same position. The receiver and transmitter are coincident, hence the distance from the object to the laser transmitter is equal to the distance from the object to the laser receiver, and the same atmospheric transmittance applies to both optical paths.

The laser beam radiance is calculated from the definition of radiance in Equation (2.19), requiring the optical power, beam area at the source and beam solid angle (divergence). In order to use this very simple equation, the Gaussian shape properties of the laser beam is discarded for two very simple uniform shapes. The laser beam angular radiance distribution is assumed to be uniform within the top-hat-shaped beam divergence profile (e.g., peak normalized divergence). The laser beam power distribution is assumed to be uniform across the area of the beam (e.g., peak normalized area). Using Equation (2.19) and these two simplifications, the radiance

can be written as

$$L_L = \frac{\Phi_L}{\Omega_L A_L}, \quad (9.21)$$

where Ω_L is the laser beam solid angle, and A_L is the laser beam cross-section area at the laser source. This simplification might not satisfy the required mathematical rigor, but it does provide order-of-magnitude radiance estimates.

From Equations (2.26) and (9.21), the irradiance on the object is then

$$\begin{aligned} E_O &= \frac{L_L A_L \tau_{OL} \cos \theta_L \cos \theta_{LO}}{R_{LO}^2} \\ &= \frac{\Phi_L \tau_{OL} \cos \theta_{LO}}{\Omega_L R_{LO}^2}, \end{aligned} \quad (9.22)$$

where it is assumed that $\cos \theta_L = 1$ because the laser beam radiates perpendicularly from the laser mirror. The uncooperative target object can have any orientation relative to the laser beam, denoted by θ_{LO} .

9.4.3 Lambertian target reflectance

The laser pulse falling onto the object is reflected by the object. Most natural surfaces have diffuse reflectance and scatter energy in all directions (a Lambertian source). The reflected laser spot on the target object has a radiance of

$$L_O = \frac{\rho E_O}{\pi} = \frac{\rho \Phi_L \tau_{LO} \cos \theta_{LO}}{\Omega_L \pi R_{LO}^2}. \quad (9.23)$$

The irradiance, caused by the reflected pulse, at the laser receiver is then, from Equation (2.26),

$$\begin{aligned} E_R &= \frac{\Phi_R}{dA_1} \\ &= \frac{L_O A_O \cos \theta_{OR} \tau_{OR}}{R_{OR}^2}, \end{aligned} \quad (9.24)$$

where A_O is the area illuminated by the laser that is visible to the sensor. Further manipulation using Equation (9.23) leads to

$$\begin{aligned} E_R &= \frac{\Phi_L \tau_{LO} \cos \theta_{LO} \rho A_O \cos \theta_{OR} \tau_{OR}}{\Omega_L \pi R_{OR}^2 R_{LO}^2} \\ &= \frac{\rho \Phi_L \tau_{LO} \cos \theta_{LO} A_O \cos \theta_{OR} \tau_{OR}}{\pi \Omega_L R_{LO}^2 R_{OR}^2}. \end{aligned} \quad (9.25)$$

By co-locating the laser transmitter and the receiver ($R_{LO}^2 = R_{OR}^2 = R$, $\tau_{LO} = \tau_{OR} = \tau_a$, and $\cos \theta_{LO} = \cos \theta_{OR}$) the expression for irradiance simplifies to

$$E_R = \frac{(\rho \cos^2 \theta_{LO} A_O) \Phi_L \tau_a^2}{\pi \Omega_L R^4}, \quad (9.26)$$

where R is the distance between the object and the rangefinder. Equation (9.26) is similar to the radar range equation. The product $\rho \cos^2 \theta_{LO} A_O$ can be regarded as the target optical cross-section. In the radar case, the optical cross section has a fixed magnitude irrespective of distance between the laser and target object. This is also true for a laser rangefinder illuminating an airborne object where there is no reflective background. If the object is observed against a terrain background, the terrain background also contributes to the reflected signal (depending on the geometry).

9.4.4 Lambertian targets against the sky

The rangefinder irradiance SNR is the ratio of signal strength [Equation (9.26)] to noise [Equation (9.20)], and is given by

$$\frac{E_R}{E_n} = \frac{\frac{\rho \Phi_L \tau_a^2 \cos^2 \theta_{LO} A_O}{\pi \Omega_L R^4}}{\frac{\sqrt{\Delta f A_d}}{D^* A_1 \tau_a}} \quad (9.27)$$

$$= \frac{\rho \Phi_L \tau_a^2 \cos^2 \theta_{OR} A_O D^* A_1 \tau_a}{\pi \Omega_L R^4 \sqrt{\Delta f A_d}}. \quad (9.28)$$

If Bouger's law is accepted for atmospheric transmittance, the transmittance can be written in terms of distance as $\tau_a = e^{-\gamma R}$. The laser flux is given by $\Phi_L \approx Q_L / t_p$, where Q_L is the pulse energy in [J], and t_p is the pulse width in [s]. The required receiver electronic noise bandwidth can be written in terms of the pulse width as $\Delta f = k_n k_f / t_p$, where k_n relates the electrical system electronic bandwidth with the noise equivalent bandwidth, and k_f relates the laser pulse width with the system electronic bandwidth (see Section 5.3.13 for both definitions).

The irradiance SNR can now be written as

$$\frac{E_R}{E_n} = \overbrace{\left(\frac{A_0 \rho \cos^2 \theta D^*}{\pi \sqrt{k_n k_f}} \right)}^{\text{no control}} \overbrace{\left(\frac{Q_L A_1 \tau_a}{\Omega_L \sqrt{t_p A_d}} \right)}^{\text{design}} \overbrace{\left(\frac{e^{-2\gamma R}}{R^4} \right)}^{\text{distance}}. \quad (9.29)$$

In Equation (9.29) there are three groups of variables:

1. Variables and constants that the designer has no control over, such as the object orientation and reflectivity, D^* , and constants.
2. Variables that the designer controls in the design process, such as laser energy, receiver aperture area, laser pulse width, and detector size.
3. Distance-related factors that the designer has little control over.

The designer can now easily determine that increased laser energy and receiver aperture improves the SNR linearly, whereas increased detector area and pulse width decrease the SNR. Contrary to intuition, a longer pulse width (i.e., a lower electronic bandwidth) decreases the SNR. Why?

9.4.5 Lambertian targets against terrain

If the laser rangefinder is viewing targets against the terrain, the laser light is reflected from the target object as well as its surrounding terrain. This implies that the real target area is not of sole importance because the terrain background also reflects the laser pulse. There are two possibilities regarding the laser transmitter and receiver beam or FOV sizes.

If the receiver FOV is larger than the transmitter beam width, the receiver sees the whole laser spot. This implies that the effective laser spot area is defined by the laser beam width by

$$\Omega_L = \frac{A_O \cos \theta_{LO}}{R^2}, \quad (9.30)$$

and hence Equation (9.26) — the irradiance at the receiver — becomes

$$E_R = \frac{\rho \Phi_L \tau_a^2 \cos \theta_{LO}}{\pi R^2}. \quad (9.31)$$

If the receiver FOV is smaller than the transmitter beam width, the receiver only views a portion of the whole laser spot. This implies that the effective laser spot area is determined by the receiver FOV as

$$\Omega_R = \frac{A_O \cos \theta_{LO}}{R^2}, \quad (9.32)$$

and hence Equation (9.26) becomes

$$E_R = \frac{\rho \Phi_L \tau_a^2 \cos \theta_{LO}}{\pi R^2} Y, \quad (9.33)$$

where $Y = \Omega_R / \Omega_L$ is the fraction of the laser spot viewed by the laser receiver FOV, and $0 \leq Y \leq 1$. When comparing Equations (9.31) and (9.33), we see that they are the same, except for the fraction Y . Equation (9.33) is the more general case because Equation (9.31) is a special case when $Y = 1$.

Table 9.4 Parameters used in rangefinder example.

Parameter	Value	Units	Parameter	Value	Units
τ_a	0.5		ρ	0.1	
A_1	2×10^{-3}	m^2	$\cos \theta$	0.5	
k_n	1		Y	1	
k_f	1		λ	1.06	μm
Ω	1×10^{-3}	sr	Q_L	0.06	J
$f/\#$	1.4		Φ	4	MW
f	0.07	m	t_p	15	ns
D^*	3×10^{11}	$\text{cm} \cdot \sqrt{\text{HzW}^{-1}}$	$d\lambda$	0.02	μm
A_d	4.6×10^{-6}	m^2			

9.4.6 Detection range

The equations in the previous section provide the signal strength obtained from a laser transmitter at a laser receiver. An estimate of the detection range can be obtained by solving the range equation

$$SNR = \frac{E_R(R)}{E_N}, \quad (9.34)$$

where SNR is the signal-to-noise ratio required to achieve detection. Solving the detection range problem means finding a value for range R in Equation (9.29) that would yield the required SNR .

9.4.7 Example calculation

Equation (9.29) is known as the laser rangefinder range equation because the range may be solved for a given set of design choices and SNR . One such solution is shown here. The values shown in Table 9.4 were used in the calculation.

Equation (9.29) was used to calculate the SNR versus range for several atmospheric conditions. The atmospheric attenuation coefficients used here³ were $\gamma = 0.17, 0.33, 0.53$, and 0.88 , corresponding to meteorological ranges of 15 km, 8 km, 5 km, and 3 km at the laser wavelength of $1.06 \mu\text{m}$. A graph indicating SNR versus range is shown in Figure 9.8. Note how strongly the atmospheric attenuation influences the range performance of the rangefinder. Figure 9.8 also shows the expected operational range as a function of detector D^* .

The background flux in the scene determines the detector current,

Table 9.5 Background radiance² at 1.06 μm , expressed as a detector D^* and resultant operating range.

Terrain	Background Radiance	Background D^*	Effective D^*	Range
Dark night	0	∞	3×10^{11}	6.64
Grass terrain	10	6×10^{11}	2.7×10^{11}	6.32
Snow terrain	100	2×10^{11}	1.67×10^{11}	5.79
Blue sky	10	6×10^{11}	2.7×10^{11}	6.32
Dark clouds	8	7×10^{11}	2.75×10^{11}	6.34
Sunlit clouds	100	1.7×10^{11}	1.48×10^{11}	5.62
	$\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \mu\text{m})$	$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$	$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$	km

which in turn determines the noise in the detector. If the sensor is operating at night, the detector noise is at a minimum. If the sensor is pointed at a bright, sunlit background, the current and hence the noise in the detector increases relative to the dark night condition. By converting the background flux noise to D^* , the range versus D^* graph can be used to predict degradation in system performance under bright sunlight conditions. Typical background radiance values are shown in Table 9.5. The background radiance values were used to calculate the current in the detector using a variation of Equation (6.16). Once the current in the detectors were known, the shot noise at the respective currents were calculated, and finally, new D^* values were calculated using Equation (5.32). These new D^* values now represent the noise performance of the sensor under the various background conditions. Once the new D^* values were known, the range performance corresponding to the different conditions were determined from the bottom graph in Figure 9.8. The operating ranges are also shown in Table 9.5.

9.4.8 Specular reflective surfaces

If the target has a specularly reflecting surface (see Figure 3.11), the surface BRDF f_r is used to calculate the reflection as a function of angle of incidence and the view direction angle (see Figure 3.12). Then

$$E_R = \frac{f_r(\theta_{LO}, \theta_{OR}) \Phi_L \tau_a^2 \cos \theta_{LO} Y}{R^2}. \quad (9.35)$$

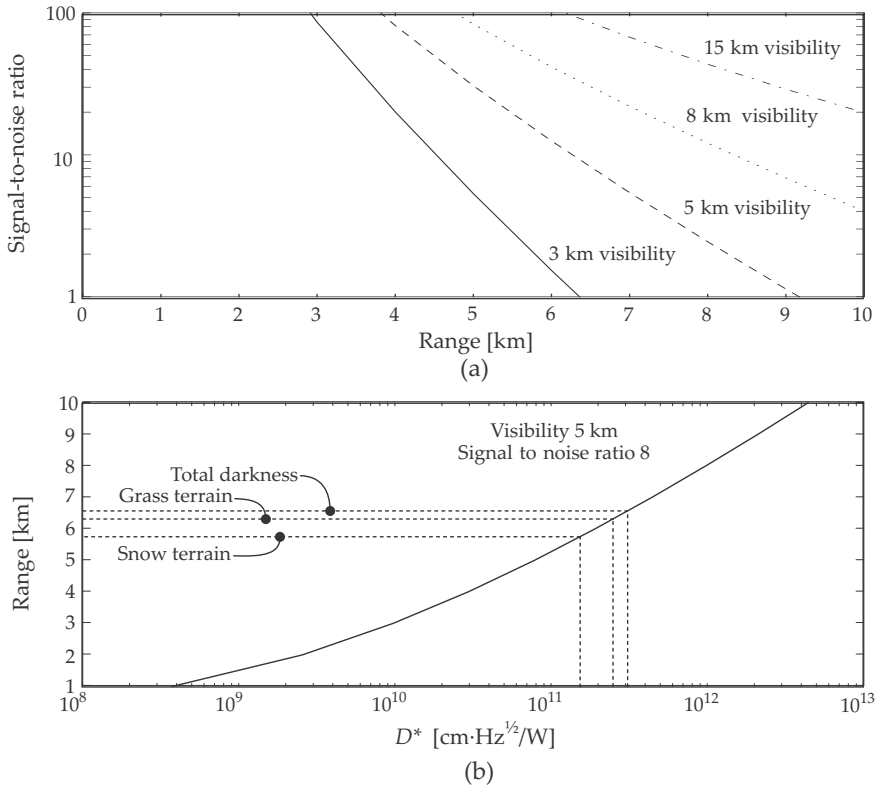


Figure 9.8 Laser-rangefinder range equation analysis: (a) SNR versus range for several atmospheric visibility values and (b) expected range as a function of detector D^* .

If the surface reflection can be modeled by the Phong equation [Equation (3.39), Figure 3.15], the irradiance at the receiver becomes

$$E_R = \left(\frac{\rho_d}{\pi} + \frac{\rho_s(n+1)\cos^n \alpha}{2\pi \cos \theta_i} \right) \left(\frac{\Phi_L \tau_a^2 \cos \theta_{LO} Y}{R^2} \right), \quad (9.36)$$

where the angles α and θ follow from the geometry given in Figure 3.12. Note that in the case of a laser rangefinder, the transmitter and receiver beams are co-axial, with the result that the \hat{I} and \hat{S} vectors align on the same axis, and hence $\theta = \theta'$ and $\alpha = 2\theta$.

Using Equation (9.20) and the values $Y = 1$, $\Delta f = 66$ MHz, $A_d = 4.6 \times 10^{-6}$ m², $D^* = 6 \times 10^{11}$ cm·√Hz/W, $A_1 = 2 \times 10^{-3}$ m², and $\tau_r = 0.5$, a sensor noise level of 3 μW is calculated. Assume further that a SNR of 5 is required for operation. The operating range is then determined by solving for R in

$$15 \times 10^{-6} = \left(\frac{\rho_{d_\theta} \cos \theta_{LO}}{\pi} + \frac{(n+1)\rho_{s_\theta} \cos^n(2\theta_{LO})}{2\pi} \right) \left(\frac{\Phi_L \exp^{-2\gamma R}}{R^2} \right), \quad (9.37)$$

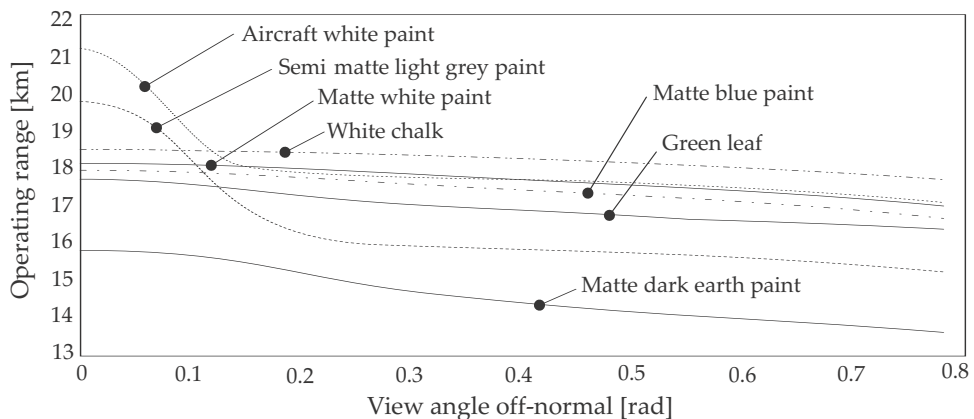


Figure 9.9 Detection range for painted and natural surfaces

where the laser power $\Phi_L = 4$ MW, and $\gamma = 1.5 \times 10^{-3}$ [1/m].

In this configuration (co-axial receiver and transmitter) strong specular reflection toward the receiver can only occur when the surface normal vector faces the rangefinder, i.e., $\theta_{LO} = 0$. By solving the equation for a few of the materials shown in Figure 8.4, the detection ranges in Figure 9.9 were obtained.

1. In Figure 9.9, note the effect of surface properties on range, and in particular, range as a function of off-normal angle.
2. Except for the white paint anomaly, Figure 9.9 shows that specular surfaces support longer detection ranges along the mirror reflection vector but lower detection ranges elsewhere. Lambertian surfaces, on the other hand, support a near-constant detection range irrespective of view angle. This is the infrared/optical equivalent of the radar geometric stealth concept. The white paint anomaly requires that this statement be closely investigated.
3. The sevenfold drop in diffuse reflectance between white chalk and matte dark earth paint, 0.77 to 0.11, resulted in a detection-range ratio of only 1.2. This indicates the relative robustness of the detection process against paint variations. The 'compression' in detection range is due to the $1/d^2$ term as well as the severe atmospheric attenuation at longer ranges.

This compression effect will be less severe in a moderate atmosphere.

4. The fivefold drop between the specular peak and diffuse reflectance for the specular paints only results in a detection range improvement of 1.2 times. The argument is the same as above.

This compression effect will be less severe in a moderate atmosphere.

5. From these results, it would appear that extraordinary attempts to reduce the laser signature by utilizing specular properties or low reflectance values are probably not worth the effort.

9.5 Case Study: Thermal Imaging Sensor Model

The sensor model described in Sections 6.5 and 6.7 is developed further to predict the performance of a thermal imaging sensor. Much of the work was already done in these two sections. Despite its simplicity, this model can be used to predict thermal camera performance in trade-off studies. Daniels⁴ covers the same topic in more detail.

The sensor is regarded as an imaging system that builds up an image by scanning the FOV with N detectors. The basic sensor configuration is shown in Figure 6.13. The total number of detector elements are scanned over the complete image by mechanical or other means. In some cases the scanning may not be optimal and some time is lost due to the scanning method. This 'lost time' is expressed in the scanning efficiency. The scanning method is not considered here.

9.5.1 Electronic parameters

Nonstaring imaging sensors construct the image by scanning a number of detectors to cover the complete FOV. This is normally done by means of mechanical movement of a prism or a mirror. In practical scanners, mechanical or optical, limitations prevent a 100% effective scan; some portion of the scan pattern can not be used for image formation. This effect is incorporated in the scan-efficiency parameter.

Consider N detector elements scanning across the image plane, contributing toward scanning the full image field. The number of electrical signal samples per frame period required to form the image is given by the total number of pixels in the image divided by the number of detector elements. The image is formed at a frame rate of F_F , hence the electronic bandwidth required to pass the detector signal (for each detector element) is defined in terms of the dwell time (time on target, integration time, or pulse width) as

$$\tau_e = \frac{\omega \eta_s N}{\Omega_r F_F \eta_a \eta_b}, \quad (9.38)$$

where Ω_r is the field of regard (the FOV covered by the sensor in one frame), ω is the pixel FOV, and η_s is the scanning efficiency. η_s is less than

unity because the scan velocity is not constant ($\eta_s = \langle v \rangle / v_{max}$) or there is a portion of the scan that is not available to form the image. The scan efficiency is effectively the ratio of useful scan period to total frame time. The scan efficiency for a staring array sensor is one.

The image fill efficiency in the a and b directions η_a and η_b allows for the situation where the detectors do not cover the total field of regard. In other words, pixel centerline spacing exceeds the pixel size. This is fairly commonplace in staring detectors, having fill factors lower than 100% fill. $\eta_a = \eta_b = 1$ implies exact filling, $\eta_a < 1$ and $\eta_b < 1$ implies under-filling, and $\eta_a > 1$ and $\eta_b > 1$ indicates overfilling.

The electronic bandwidth required to pass the signal is given by

$$f_{-3 \text{ dB}} = \frac{k_f}{\tau_e} = \frac{k_f \Omega_r F_F \eta_a \eta_b}{\omega \eta_s N}, \quad (9.39)$$

where k_f is the time-bandwidth product (see Section 5.3.14). The noise equivalent bandwidth can be derived from Equation (9.39) as

$$\Delta f = \frac{k_n k_f \Omega_r F_F \eta_a \eta_b}{\omega \eta_s N}, \quad (9.40)$$

where k_n is the ratio of noise equivalent bandwidth to -3 dB bandwidth (refer to Section 5.3.13).

9.5.2 Noise expressed as D^*

All of the noise sources in the sensor can be combined into one single number. It is convenient to express this number in terms of the detector D^* , which can be derived from Equations (5.30) and (5.26).

9.5.3 Noise in the entrance aperture

The NEE in the sensor's entrance aperture is given by

$$NEE_S = \frac{\sqrt{\Delta f A_d}}{k_s D^*_{\text{eff}} A_s \tau_s}, \quad (9.41)$$

where NEE_S is the inband NEE, D^* is derived from Equations (5.30) and (5.26), Δf is the noise equivalent bandwidth, A_d is the area of the detector, A_s is the area of the sensor's entrance pupil, and τ_s is the effective transmittance of the sensor. The optical PSF constant k_s is the fraction of energy from a point source falling onto a single detector element. It therefore represents the sensor's capability to gather energy from a point source. It is assumed that each detector element has uniform responsivity over its area.

By mathematical manipulation the NEE is developed as

$$NEE = \frac{\sqrt{\Delta f A_d}}{k_s D_{\text{eff}}^* A_s \tau_s} \quad (9.42)$$

$$= \frac{\sqrt{\eta_a \eta_b k_n k_f \Omega_r F_F a b} 4(f/\#)^2}{k_s \sqrt{\omega \eta_s \bar{N}} D_{\text{eff}}^* P \pi f^2 \tau_s} \quad (9.43)$$

$$= \frac{\sqrt{\eta_a \eta_b k_n k_f \Omega_r F_F \omega} 4(f/\#)^2}{k_s \sqrt{a b \eta_s \bar{N}} D_{\text{eff}}^* P \pi \tau_s} \quad (9.44)$$

$$= \frac{\sqrt{\eta_a \eta_b k_n k_f \Omega_r F_F} 4(f/\#)^2}{k_s \sqrt{\eta_s \bar{N}} D_{\text{eff}}^* P \pi f \tau_s}. \quad (9.45)$$

If a single-pole Butterworth filter is employed, the constant $k_n = \pi/2$. Assume that $k_f = \pi/2$. The (conservative) equation for NEE then becomes

$$NEE = \frac{\sqrt{\eta_a \eta_b \Omega_r F_F} 2(f/\#)^2}{k_s \sqrt{\eta_s \bar{N}} D_{\text{eff}}^* P f \tau_s}. \quad (9.46)$$

9.5.4 Noise in the object plane

Section 6.7 describes a mechanism for transforming noise to different planes in the system. The noise can be referred to the object plane *for extended sources* by noting that $k_s = 1$, and

$$NEE = \frac{NEL A_0 \cos \theta + 0}{R^2} = NEL \omega = \frac{NEM \omega}{\pi}. \quad (9.47)$$

where the projected solid angle should be used for ω because the thermal camera senses Lambertian sources. However, for the small pixel FOV generally used, the projected and geometrical solid angles are numerically equal.

Using Equation (9.45), the noise equivalent exitance (NEM) is therefore given by

$$NEM = \frac{\pi NEE}{\omega} = \frac{\sqrt{\eta_a \eta_b k_n k_f \Omega_r F_F} 4(f/\#)^2}{\alpha \beta \sqrt{\eta_s \bar{N}} D_{\text{eff}}^* P f \tau_s}. \quad (9.48)$$

The noise equivalent temperature difference (NETD) is the temperature in the source that causes the same signal as the noise in the sensor. It is given by

$$\frac{NEM}{NETD} = \frac{dM}{dT}$$

$$\begin{aligned}
 NETD &= \frac{\pi NEEf^2}{ab \frac{dM}{dT}} \\
 &= \frac{\sqrt{\eta_a \eta_b k_n k_f \Omega_r F_F} 4(f/\#)^2}{\alpha \beta \sqrt{\eta_s N} D_{\text{eff}}^* P f \tau_s \frac{dM}{dT}}, \quad (9.49)
 \end{aligned}$$

where

$$\frac{dM}{dT} = \int_0^\infty \epsilon_\lambda \tau_{a\lambda} \mathcal{S}_\lambda \frac{dM_\lambda(T_t)}{dT} d\lambda \quad (9.50)$$

is the derivative of the source exitance with respect to the source temperature; in other words, the rate at which the source exitance changes for a given change in source temperature. This derivative is required to transform a small change in exitance into a corresponding small change in temperature. It is required that the background temperature be specified with the NETD value because the NETD value depends on the background temperature against which it is measured. NETD is generally used to describe the sensitivity of thermal imaging systems. NETD is only defined for extended sources, that is, sources that are larger than the sensor pixel FOV. If the source is smaller than the FOV, any attempt to describe the NETD is incorrect.

By simplification, the equation for noise equivalent temperature difference, Equation (9.49), can be reduced to the form of Equation (5.35) in Lloyd's classic book:⁵

$$NETD = \frac{2\sqrt{\Delta f}}{\sqrt{\alpha \beta \eta D^* \tau_s D_s} \frac{dM}{dT}}. \quad (9.51)$$

9.5.5 Example calculation

Consider now the application of Equation (9.49) for the performance prediction of a thermal camera. Two spectral bands are investigated: 3–5.5 μm and 8–12 μm . In addition, the relative value of using a large number of detector elements must be investigated. In particular, we investigate cameras with $N = 1$, 256, and 256×256 detector elements. For the single-element detector case it is assumed that the image is formed by a two-dimensional scanner sweeping the single detector to form the complete image. For the 256-element detector case it is assumed that the image is formed by a one-dimensional scanner sweeping a vector of detectors to form the complete image. For the 256×256 element detector case it is assumed that the image is formed by a staring array sensor with no scanning. The design parameters considered for this analysis are shown in Table 9.6.

Table 9.6 Thermal camera analysis parameters.

Parameter	Value	Units	Parameter	Value	Units
k_n	1.15		$f/\#$	1.2	
k_f	1		P	1	
η_s	0.6/ 0.9/ 1		f	0.1956	m
η_a	1		η_b	1	
pixels	256×256		$a = b$	40	μm
Ω_r	3×3	deg	ω	0.2×0.2	mrad^2
F_F	25	Hz	τ_s	0.8	
3–5 μm parameters					
dM/dT	≈ 0.37	$\text{W}/(\text{m}^2 \cdot \text{K})$	D_{eff}^*	6×10^{10}	$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$
8–12 μm parameters					
dM/dT	≈ 2	$\text{W}/(\text{m}^2 \cdot \text{K})$	D_{eff}^*	2×10^{10}	$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$

Table 9.7 Thermal camera performance for three detector configurations.

Detectors	τ_e	Δf	NETD 3–5 μm	NETD 8–12 μm
1	366 ns	3.14 MHz	1.44 K	0.80 K
256	141 μs	8.2 kHz	0.073 K	0.041 K
256×256	40 ms	28.7 Hz	0.0043K	0.0024 K

For some parameters three values are listed. In these cases they correspond to the detector-element choices of 1, 256, and 256×256 . Note that not all of the values assumed above are realistic in practice. It is assumed that the detector arrays will have perfect responsivity uniformity. This is not practical or possible in real life. However, from a design comparison perspective, these values are accepted. The electronic noise bandwidth, detector dwell time, and NETD values for the three detector choices are shown in Table 9.7.

9.6 Case Study: Atmosphere and Thermal Camera Sensitivity

One of the key performance parameters of a thermal imager is its sensitivity expressed as the ‘noise equivalent temperature difference’ or NETD [Equation (9.49)]. This parameter indicates the noise level of the imager expressed as a temperature contrast at the target. The NETD represents the smallest temperature difference that can be measured by electronic instru-

ments. The minimum resolvable temperature (MRT) by a human observer is less than the NETD because the human eye and brain temporally integrates and interprets the image.

In the evaluation of thermal cameras, the target signature is commonly stated as a temperature difference with respect to the scene or background temperature, e.g., $\Delta T = 2$ K. When so specified, the sensitivity (NETD) of the thermal imager depends on the scene temperature T_t [refer to Equation (9.50)]. The colder the temperature, the less sensitive the camera becomes (higher NETD means poorer performance). Equation (9.50) allows for atmospheric transmittance correction; this can be significant over longer ranges. This section extends the strict notion of NETD as a laboratory measurement at zero target distance (no atmosphere) to account for atmospheric attenuation by inclusion of the atmospheric attenuation τ_a . The combined use of Equations (9.49) and (9.50) allows the calculation of the target signal required to overcome the atmospheric attenuation and then to provide a signal equal to the noise. For convenience, call this the 'noise equivalent target contrast' (NETC).

As a practical application of the imaging sensor model, evaluate the relative performance of a 3–5- μm imager versus an 8–12- μm imager when effected by water vapor in the atmosphere. The effect of atmospheric water-vapor attenuation is discussed in Section 4.6.8. This section evaluates the NETC performance for the following cases:

1. Practically available technology in the year 2000: staring array for 3–5- μm imagers and linear scanned arrays for 8–12- μm imagers. All other parameters are the same. The number of detector elements in the linear vector is equal to the square root of the number of elements in the staring array.
2. Practically available technology in the year 2012: Two identical imager staring array configurations using the same optics, detector configuration, and other design parameters. In this case, the performance of the spectral band is tested in absolute terms.

Figure 9.10 shows the comparison for three levels of humidity (50%, 75%, and 95%), for atmospheric conditions ranging from -20°C to $+50^\circ\text{C}$. The background temperature is assumed to be the same as the atmospheric temperature. Four distances are considered: 0 km, 2.5 km, 5 km, and 10 km. The noise equivalent target contrast is calculated for all of these conditions; see Figure 9.10. The curves show the NETC versus scene/background temperature. Observe that the 8–12- μm imager performance degrades rapidly at higher temperatures and longer ranges. The NETC for a

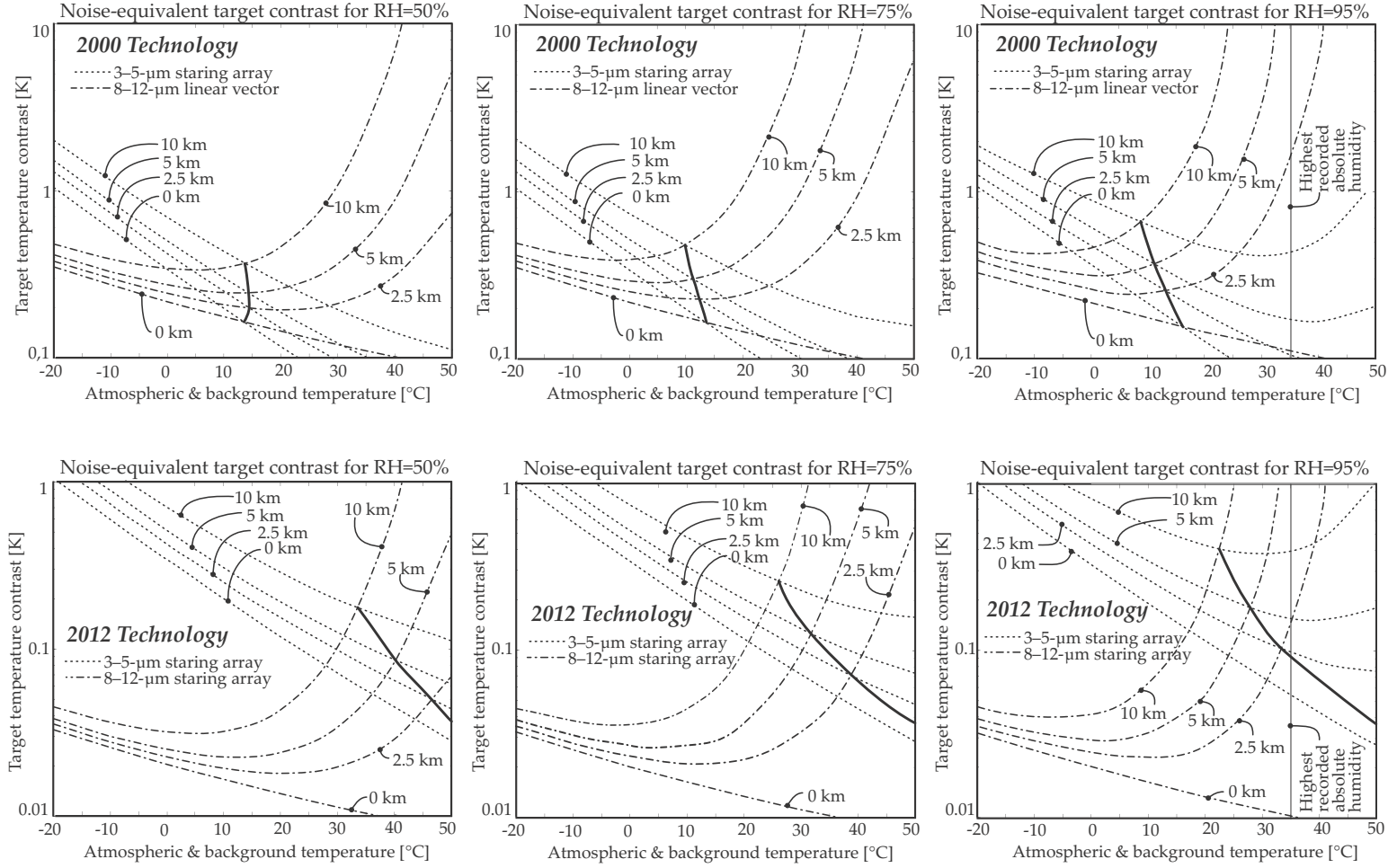


Figure 9.10 Noise equivalent target contrast required for different 3–5- μm imagers and 8–12- μm imagers. Higher NETC means poorer performance.

3–5- μm imager increases with decreasing temperature and shows a slight increase for higher relative humidities but not as severely as the 8–12- μm imager.

The graphs in Figure 9.10 show, in the thick line, the cross-over temperature where the two systems perform equally. Below this cross-over temperature, the 8–12- μm thermal imager performs better and above the cross-over; the 3–5- μm imager performs better.

The bottom graphs in Figure 9.10 indicate that, for the same detector/scanner configuration, the 8–12- μm imager outperforms the 3–5- μm imager when observing objects in a cooler environment. The top graphs compare the staring array 3–5- μm imager with the linear scanning 8–12- μm imager. It is clear that the staring 3–5- μm imager performs on par with the scanning 8–12- μm imager, at moderate and higher temperatures.

In the case of year-2000 technology, NETC compares equal for 8–12- μm imagers and 3–5- μm imagers over a very broad band of moderate climates, and benefits only occur at extreme climatic conditions.

In the case of year-2012 technology, the 8–12- μm imager outperforms the 3–5- μm imager for low temperature and moderate climatic conditions, whereas the 3–5- μm imager remains the imager of choice only for extremely high humidity conditions.

9.7 Case Study: Infrared Sensor Radiometry

9.7.1 Flux on the detector

The physical components normally found in an infrared sensor with a cooled detector are shown in Figure 9.11. In sensors without cooling, the cooler and associated components will not be present.

The detector element is mounted in a thermally insulated thermos flask, called a dewar. The dewar serves to protect the detector, but also provides thermal insulation to maintain the detector temperature at some low operating temperature. In order to prevent thermally generated charges in the detector, the detector is cooled down by one of several different cooler devices.^{6,7} Different detector types operate at temperatures ranging from several kelvin to 200 K, depending on the detector's material type and spectral range. The detector element is mounted on the front of a 'cold finger,' the whole tip of which is at the cold temperature.

Shot noise in the detector depends on the photon-flux-induced current in the detector. Lower noise can be achieved by reducing the background

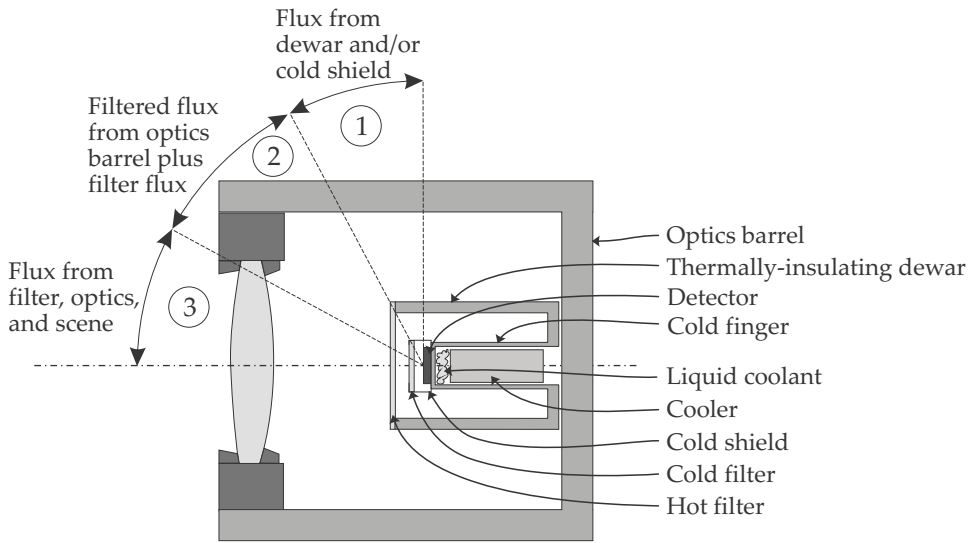


Figure 9.11 Infrared sensor layout.

flux on the detector. In order to reduce the unwanted flux on the detector, a cold shield^{8–11} (cold screen or cold cone) is constructed around the detector, mounted on the cold finger (see Figure 5.21). Because this cold finger and shield are at the same low temperature as the detector, the thermal radiation from the cold shield is significantly less than the radiance from the sensor components at room temperature.

The sensor may also employ an optically selective filter, mounted in front of the detector. This filter only transmits flux in the spectral band required by the sensor's application. Flux outside the transmittance passband is attenuated and never reaches the detector. Note, however, that the filter also emits flux because the filter is a thermal radiator. Thus, in the passband, the filter transmits flux from the scene, whereas in the stopbands the filter radiates with emissivity $\epsilon = \alpha = 1 - \tau - \rho$, from Equation (2.3.4) and Kirchhoff's law (Section 3.2.1). In Figure 9.11 the filter is shown as the front dewar window, but it can be located anywhere in the optical path. The filter can be cooled down to reduce its radiated flux. If the filter is mounted on the cold finger, it is called a 'cold filter.'

Figure 9.11 depicts the different radiance zones in the sensor. To the first approximation, the zones are rotationally symmetric even though the figure depicts these as linear angles. Also, the zones are shown to emanate from the center of the detector, but any real detector has a finite size with slightly different zone shapes from every small part on the detector. The cold shield geometry can become quite complex in a detailed analysis.

In zone 1 the flux on the detector originates on the walls of the dewar (sensor temperature) and/or cold shield walls (detector temperature). Clearly, the design objective will be to increase the cold-shield solid angle so as to decrease the radiation from the hot dewar walls and optics barrel without reducing the signal flux.

In zone 2 the flux on the detector originates on the inside of the sensor, i.e., the optical barrel and mounting rings. This flux could be thermally emitted or external flux reflected from the barrel. The internal flux is filtered by the hot/cold filter. If the filter is hot, the flux in the stopbands would be of the same magnitude as the internally self-emitted flux (suppressing the reflected flux). If the filter is cold, the filter flux in the stopbands can be small.

In zone 3 the flux on the detector is the sum of the scene flux, the optics flux, and the filter flux. In the event of a sensor with hot optics and filter, together with a cold scene, the scene flux can be considerably less than the sensor fluxes.

A key strategy in improving a sensor's sensitivity is therefore to cool down the detector environment and filter to reduce the background flux in zones 1 and 2. A cold filter will also reduce the filter radiated flux $L(T_f)(1 - \tau - \rho)$ in zone 3, which can be significant if the filter has a narrow passband. Optics flux can be minimized by keeping the optics cool, but more importantly, selecting materials with low emissivity in the sensor spectral band. For a discussion of the effect of hot optics, see Section 9.8.2.

The radiometry in an imaging system is described in the camera equation,^{12,13} which will be further developed in this section. The derivation will cover the primary scene radiance as well as (some of) the radiance sources in the sensor itself. In this analysis, the ideal thin-lens paraxial approximation is made. The linear angles in the sectional diagrams should be viewed as rotationally symmetric solid angles, e.g., zone 3 in Figure 9.11 is a conical solid angle. It is also assumed that system throughput is the same at all field angles, i.e., the on-axis marginal ray cone solid angle has the same value as the off-axis marginal ray cone solid angle, as in Figure 9.12.

9.7.2 Focused optics

Figure 9.12 shows the primary flux sources in an imaging sensor. The contributing source radiance values are the scene focused on the detector (L_{scene}), the optics and window (L_{optics}), the filter (L_{filter}), the optics barrel and inside of the sensor (L_{barrel}), and the detector cold shield ($L_{\text{cold shield}}$).

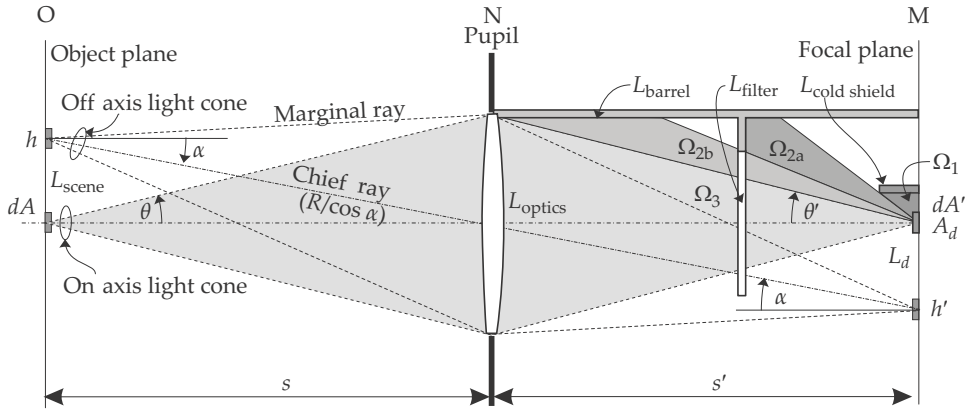


Figure 9.12 Radiometry in an imaging system.

The flux falling in on the detector is then given by

$$\Phi_{\text{det}\lambda} = \frac{A_d A_o \cos \alpha \cos \alpha \tau_{o\lambda} \tau_{f\lambda} L_{\text{scene}\lambda}}{(s' / \cos \alpha)^2} + A_d \Omega_3 (\tau_{f\lambda} L_{\text{optics}\lambda} + L_{\text{filter}\lambda}) + A_d (\Omega_{2b} \tau_{f\lambda} L_{\text{barrel}\lambda} + \Omega_{2a} L_{\text{barrel}\lambda} + \Omega_1 L_{\text{cold shield}\lambda}), \quad (9.52)$$

where A_d is the detector area, A_o is the optics exit pupil area, α is the off-axis angle to the object, $\tau_{o\lambda}$ is the optics transmittance, $\tau_{f\lambda}$ is the filter transmittance, $L_{\text{scene}\lambda}$ is the scene radiance [Equation (8.1)], $L_{\text{optics}\lambda}$ is the optics radiance, $L_{\text{filter}\lambda}$ is the filter radiance, $L_{\text{barrel}\lambda}$ is the optics barrel radiance, and $L_{\text{cold shield}\lambda}$ is the detector cold shield radiance. The solid angles Ω_1 , Ω_{2a} , Ω_{2b} , and Ω_3 are defined in Figure 9.12. From Equation (2.3), $\epsilon_{o\lambda} = 1 - \rho_{o\lambda} - \tau_{o\lambda}$, and $\epsilon_{f\lambda} = 1 - \rho_{f\lambda} - \tau_{f\lambda}$. Note that $A_o/s' = \Omega_3 = \pi \sin^2 \theta'$. Equation (9.52) assumes that the optical system has no vignetting or central obscuration. Analysis of any real system would require that these factors be taken into account. Not shown in any of the derivations here is the effect of stray light entering from outside of the ray cone. The source can be outside the sensor or inside the sensor (e.g., hot rotating parts). The stray light is normally associated by one or more reflections from the optical barrel or even optical element surfaces. Stray light is often suppressed by appropriate baffle design, but some stray effects may remain. Using Equation (2.12) for Ω_3 , the detector flux is then

$$\begin{aligned} \Phi_{\text{det}\lambda} = & A_d \pi \sin^2 \theta' \tau_{o\lambda} \tau_{f\lambda} L_{\text{scene}\lambda} \cos^4 \alpha \\ & + A_d \pi \sin^2 \theta' \tau_{f\lambda} (1 - \rho_{o\lambda} - \tau_{o\lambda}) L_{\lambda}(T_{\text{optics}}) \\ & + A_d \pi \sin^2 \theta' (1 - \rho_{f\lambda} - \tau_{f\lambda}) L_{\lambda}(T_{\text{filter}}) \end{aligned}$$

$$\begin{aligned}
& + A_d \Omega_{2b} \tau_{f\lambda} \epsilon_{\text{barrel}\lambda} L_\lambda(T_{\text{barrel}}) \\
& + A_d \Omega_{2a} \epsilon_{\text{barrel}\lambda} L_\lambda(T_{\text{barrel}}) \\
& + A_d \Omega_1 \epsilon_{\text{cold shield}\lambda} L_\lambda(T_{\text{cold shield}}),
\end{aligned} \tag{9.53}$$

where $L_\lambda(T)$ is the spectral Planck-law radiation at temperature T , T_{optics} is the optics (window and optical elements) temperature, T_{filter} is the filter temperature, T_{barrel} is the optics barrel temperature, $T_{\text{cold shield}}$ is the cold shield temperature, $\epsilon_{\text{barrel}\lambda}$ is the barrel spectral emissivity, $\epsilon_{\text{cold shield}\lambda}$ is the cold shield spectral emissivity, and θ' is the field angle (maximum inclination of the marginal ray). The value $\sin \theta'$ is known as the numerical aperture (NA) of the lens (see Section 6.3.3).

For the ideal lens (no aberrations, perfectly flat image and object planes, and obedience to the Abbe sine condition), and at infinite conjugates $\sin \theta' = NA = D/2f = 1/2F_\#$. The first term then becomes

$$\begin{aligned}
\Phi_{\text{det scene } \lambda} &= A_d \pi (NA)^2 \tau_{o\lambda} \tau_{f\lambda} L_{\text{scene}\lambda} \cos^4 \alpha \\
&= \frac{A_d \pi}{4F_\#^2} \tau_{o\lambda} \tau_{f\lambda} L_{\text{scene}\lambda} \cos^4 \alpha.
\end{aligned} \tag{9.54}$$

If the object is not at infinity (finite conjugates), the optics image the object onto the focal plane with a given magnification m . Combining the simple lens equations, Equations (6.1), (6.2), and (6.3), with Equation (9.54), it is found that $s' = f(1 + |m|)$. Applying that to the definition of θ it follows that (for paraxial optics)

$$\sin \theta = \frac{D}{2(1 + |m|)f} = \frac{1}{2F_\#(1 + |m|)}, \tag{9.55}$$

leading to the new formulation for flux on the detector:

$$\Phi_{\text{det scene } \lambda} = \frac{\pi K_C K_N(\alpha) A_d \tau_{o\lambda} \tau_{f\lambda} L_{\text{scene}\lambda} \cos^4 \alpha}{4F_\#^2 (1 + |m|)^2}, \tag{9.56}$$

where two new factors are introduced: K_C accounts for central obscuration (if present), and $K_N(\alpha)$ accounts for vignetting.¹³ Note that the angle under consideration in the \cos^4 term is the object field angle, and not the cold shield angle, which is a function of the numerical aperture (f -number cone) of the cold shield. Unless the cold shield obscures the optical ray cone, the cold shield numerical aperture does not come into consideration at all in the \cos^4 effect.

The cold shield efficiency is the ratio of scene flux to total flux onto the detector [Equations (9.52) and (9.56)],

$$\eta_{\text{cold shield}} = \frac{\int \Phi_{\text{det scene } \lambda} d\lambda}{\int \Phi_{\text{det}\lambda} d\lambda}. \tag{9.57}$$

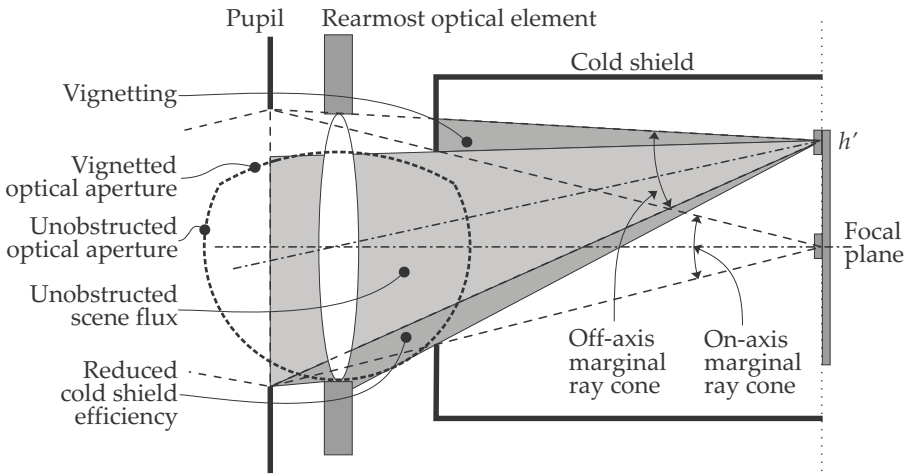


Figure 9.13 Reduced cold shield efficiency and vignetting in practical designs.

The objective of cold shield design is to achieve as high an efficiency as possible. It is however very difficult to reach high efficiency values for fast optics (low f -number) and large detectors. Ideally, the cold shield should screen or shield all of the sensor internals behind the exit pupil; the cold shield aperture should coincide with the exit pupil. If the cold shield coincides with the exit pupil, it is called a cold stop. Practical cold shield design requires the cold shield numerical aperture to be slightly larger than the optics' in order to allow for the off-axis imaging rays. Consider the picture in Figure 9.13 where the cold shield numerical aperture exactly equals the optics' numerical aperture (i.e., the same marginal ray), but the cold shield aperture is displaced from the exit pupil. The ray cone at nonzero field angles will be partially vignetted (loss of scene flux), and in addition, some internal sensor parts will be observed (reduced cold shield efficiency). If the cold shield numerical aperture is increased to reduce vignetting, the cold shield efficiency is also reduced.

9.7.3 Out-of-focus optics

In Figure 9.14, consider a small elemental area dA_i in the object plane imaging onto a small elemental area dA' in the focal (image) plane. The conical solid angle defined by the marginal rays (solid of revolution) contains all of the flux flowing from dA_i to dA' . No flux outside this solid angle contributes to the flux flow. Further, consider a small portion of the solid angle as shown in the dark shaded area in the top figure of Figure 9.14. All of the flux flowing from dA_i to dA' , passing through dA_o in the plane O_o ,

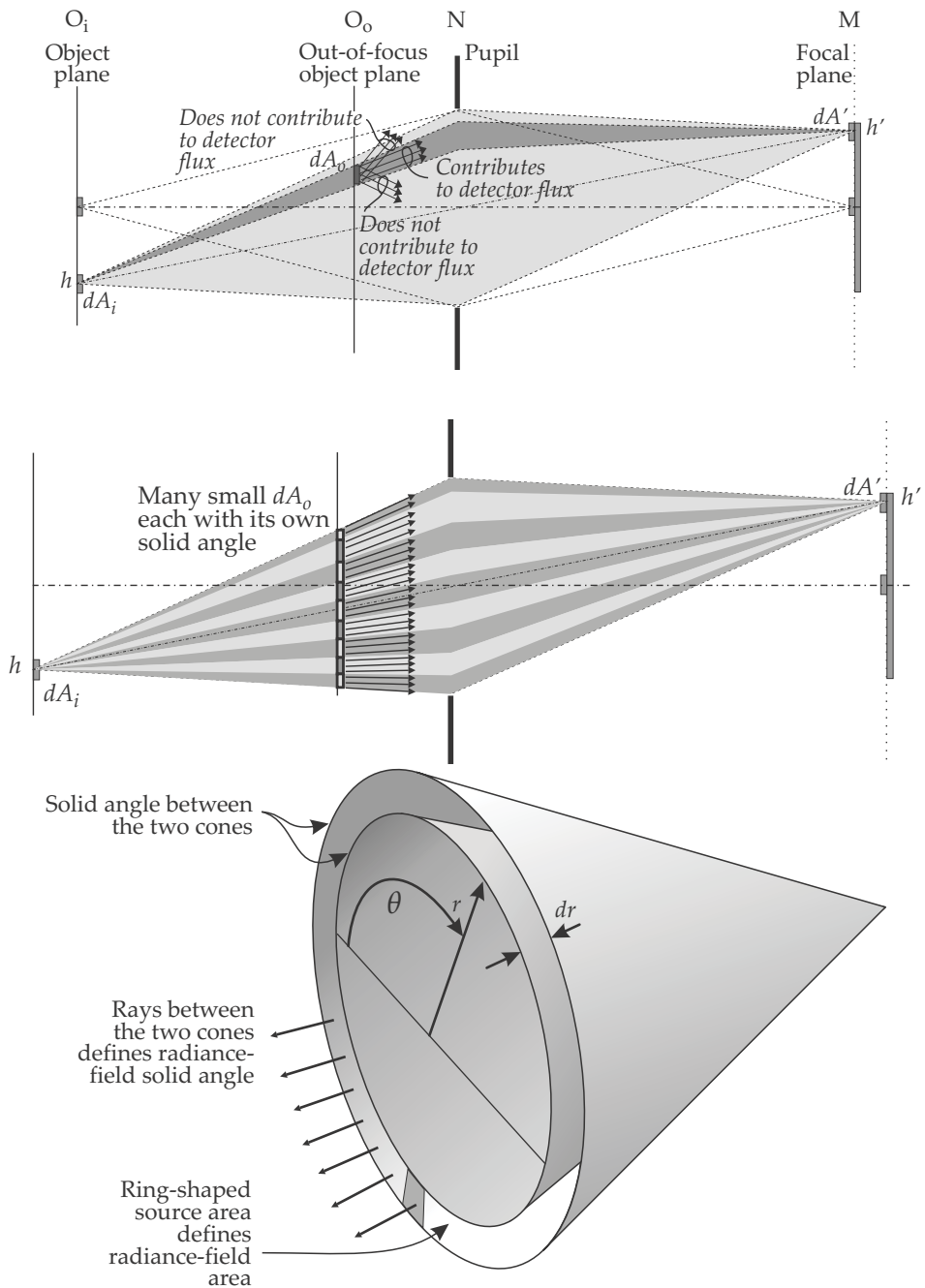


Figure 9.14 Ideal optics; out-of-focus object radiance.

has to flow along the shaded solid angle indicated in the figure. Any flux flowing through the area dA_o but outside the dark shaded solid angle will not contribute to the flux flow between dA_i and dA' . This is shown more explicitly in the bottom two pictures in Figure 9.14.

Suppose that an opaque source with uniform radiance is located in plane O_o . The optics located in N are not concerned with where the flux emanates, from the plane O_i or O_o . The optics dutifully focus the rays according to their image forming design, along the same ray paths in both cases. The focal-plane irradiance of an out-of-focus source dA_o of uniform radiance provides exactly the same flux on the detector element as would the in-focus source dA_i with the same radiance. This observation is a re-statement of the principle of radiance conservation: the location of the source is not important, the spatial properties of the radiance field determines the flux in the focal plane. The properties (including location) of the source *are* important to create the radiance field; the field cannot exist without the source in its precise location. However, once created, the radiance field 'carries' through space with no further dependence on the source.

Careful study of Figure 9.14 also indicates that radiance is defined by a matched set of areas and solid angle directions. The small area dA_i in the object plane, with a full conical solid angle uniformly filled with rays, provides the same flux on the detector as a large number of small areas in the plane O_o but with narrow conical sections associated with each small area.

The practical implication of this observation is that large, uniform sources do not have to be 'in focus.' The requirement for the object location at a particular plane of focus is only a requirement if the source radiance is not uniform, and this pattern must be imaged accurately onto the focal plane, i.e., if a sharp image is required.

9.8 Case Study: Bunsen Burner Flame Characterization

This case study provides an overview of a simple approach to flame signature characterization. The process shown here can be used as the basis for a more-advanced analysis procedure. Not all of the data and results are shown here; these are available on the pyradi website.¹ See also an alternative investigation.¹⁴

A laboratory bunsen burner was characterized with the objective to determine its area, temperature, and emissivity (Section 8.4). The burner was adjusted for no premix of air with the gas to obtain a very long, yellow

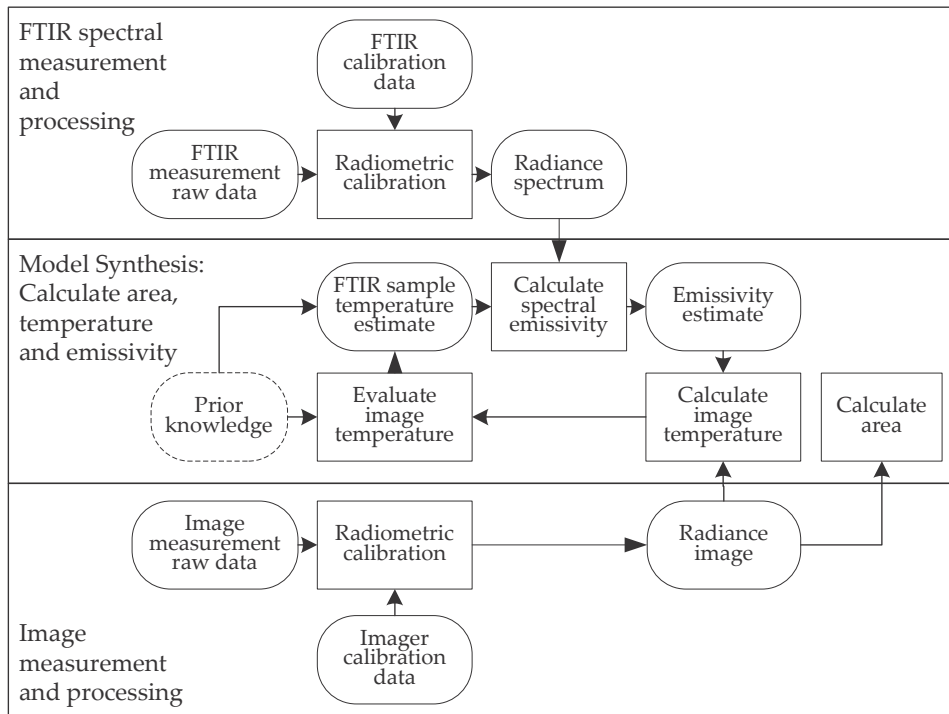


Figure 9.15 Flame data analysis workflow.

flame. At the exit of the burner nozzle, the flame is rich in butane/propane and less rich in oxygen. The gas mixes with the air at the ‘outside’ and top of the flame, burning away the rich gas concentration at the center of the flame.

9.8.1 Data analysis workflow

The data analysis workflow is shown in Figure 9.15. The process starts with the two sets (flame and reference) of measurements. In each case the instrument calibration data is used to calculate radiance values (a radiance spectrum and a radiance image). Because (1) the flame only partially fills the FTIR spectrometer FOV, but (2) the calibration data applies to a fully filled FOV, the spectral radiance measurement is ‘scaled’ with an unknown factor.

The temperature and spectral emissivity jointly result in the observed radiance values. The magnitudes of neither temperature nor spectral emissivity are known. This procedure investigates sets of temperature and emissivity values (Section 8.4) that would solve the measurement equation. One set — the one that best matches physical reality — is finally

selected.

Starting from prior knowledge (open literature, past experience, or physics principles), select a temperature and then calculate spectral emissivity from the measured spectral radiance and Planck's law:

$$\epsilon_{\lambda m} = \frac{L_{m\lambda}}{L_{bb\lambda}(T_m)}, \quad (9.58)$$

where $L_{m\lambda}$ is the measured radiance, and $L_{bb\lambda}(T_m)$ is the blackbody radiance at the estimated temperature.

Use the radiance image obtained from the imaging instrument and the spectral emissivity estimate $\epsilon_{m\lambda}$ obtained above to calculate a temperature image (for each pixel in the radiance image) by solving for temperature in

$$L_{\text{image}} = \int_0^\infty L_{bb\lambda}(T_m) \epsilon_{\lambda m} \tau_{a\lambda} \mathcal{S}_\lambda d\lambda, \quad (9.59)$$

where L_{image} is the wideband radiance value measured by the imaging camera, $\epsilon_{\lambda m}$ is calculated above, $\tau_{a\lambda}$ is the spectral atmospheric transmittance, and \mathcal{S}_λ is the imaging sensor spectral response. This process provides an image where each pixel in the image represents temperature. The temperature map is critically evaluated to determine if the predicted temperatures are acceptable (judgement call required). If the temperatures are not acceptable, the FTIR sample temperature estimate is adjusted, and the process is repeated.

This analysis assumes that the spectral emissivity will not vary significantly for different flame temperatures — this assumption is only warranted if the temperature spread is not too wide. This analysis also ignores the effect of atmospheric transmission, i.e., $\tau_{a\lambda} = 1$ in Equation (9.59), on the assumption that the path length in the laboratory was short. However, the atmospheric attenuation in the CO₂ absorption band is severe, even over these short distances.

The model developed here serves to convey the principles involved. In practice, these principles will be used to develop a model that accounts for the spatial variation across the area of the flame (texture). In some cases the model might even account for temporal variations in the texture and flame radiance.

9.8.2 Instrument calibration

The imaging radiometers and spectrometer were calibrated (or characterized) against laboratory blackbody sources. In most cases the calibration is an elaborate process covering several instrument settings, filters, and environmental conditions, but it can be summarized as follows. Calibration

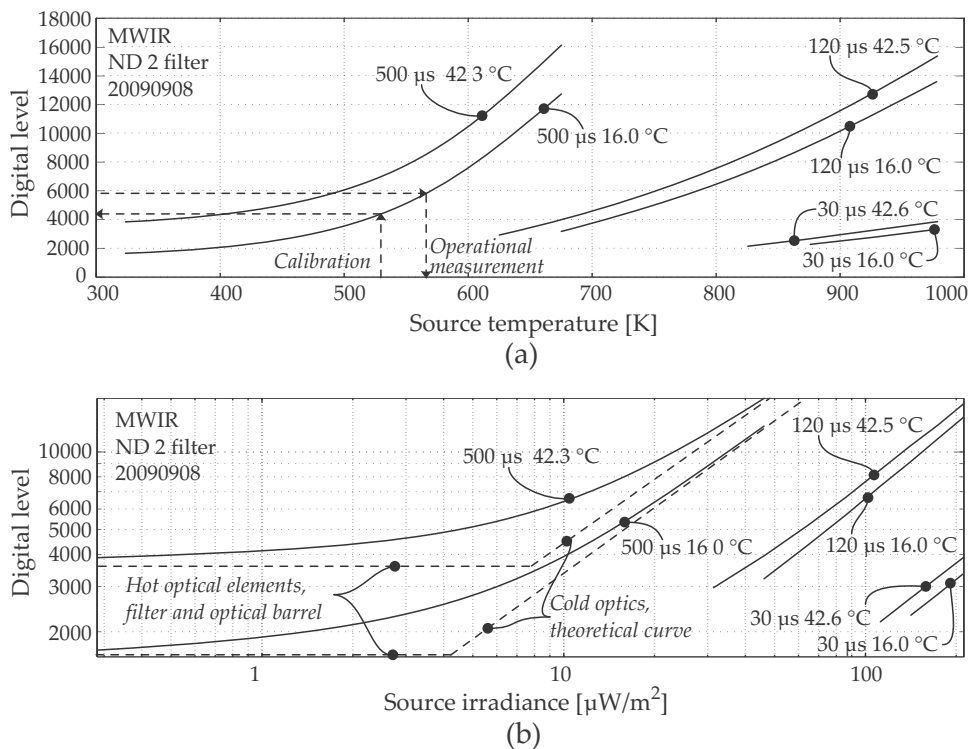


Figure 9.16 MWIR imaging radiometer calibration curves for (a) digital level vs. blackbody temperature, and (b) digital level vs. irradiance.

entails measuring the instrument output (voltage or digital levels) versus source temperature for a series of source temperatures (ideally over the full dynamic range of the instrument). The set of measurements form a temperature calibration curve (Figure 9.16). When using the instrument, the curve is read 'backward' such that a given instrument voltage returns the appropriate source temperature. If the test object's emissivity is the same as the laboratory source emissivity, the object's temperature will be equal to the source temperature. Section 8.5 elaborates in more detail on the effect of the object surface emissivity and other flux contributions during such a measurement.

Accurate measurement work requires that the spectral sensor response [Equation (6.15)] must be known. In this case, the calibration source temperature, together with the spectral sensor response, can be used to calculate the inband source radiance [using Equation (6.17)]. A new curve is now constructed to relate signal voltage with inband radiance for an extended source completely filling the pixels. This curve can now be used to determine the radiance field incident on the instrument during a measurement. This radiance value lends itself more readily to the analysis

described in Section 6.6 and Chapter 8. It is also the basis for the analysis described in this section.

Figure 9.16 shows typical calibration data. The top curve relates source temperature and instrument voltage (digital level). The bottom curve relates source irradiance on the sensor entrance aperture to instrument voltage. For an extended source, the source apparent radiance is related to irradiance by $E = L\omega$, where ω is the pixel FOV. In this case the instrument was calibrated for three different gain settings (integration times of 30, 120, and 500 μs) and two different sensor internal temperature conditions (16 and 42.6 $^{\circ}\text{C}$). Observe the importance of calibration at different internal temperatures: the curves (for the same instrument and settings) show significantly different responses. A good strategy to allow for temperature changes in the sensor is to measure the calibration curve at several different internal temperatures and then interpolate between these, according to the actual instrument temperature during the measurement.

The effect of hot optics is also shown in Figure 9.16. It is evident that the hot optics flux sets an asymptotically lower measurable flux limit (called the ‘floor’). When measuring low-temperature test samples, a small variation in internal temperature shifts the floor up or down, playing havoc with calibration. In this particular instrument setting, an ND2 neutral density filter (0.01 transmittance) was used. Much of the hot optics flux emanates from this filter. So in all fairness, there is little point in using an ND2 filter when measuring a low-temperature target. The situation does arise, however, if there is a requirement to measure both hot and cold test samples without changing filters.

9.8.3 Measurements

Measurements were made with imaging cameras operating in the 3–5- μm MWIR and 7–11- μm LWIR spectral ranges, a Fourier transform infrared (FTIR) spectrometer operating in the 2.5–5.5- μm spectral range, and with a thermocouple.

The temperature in various locations in the flame was determined with a thermocouple measurement. This thermocouple reading was found to be very difficult because the flame temperatures appeared to vary quite significantly, and the readings were considerably lower than expected.

As part of the measurement, a set of ‘reference’ measurements were also made of laboratory blackbody sources at known temperatures. These measurements serve to confirm instrument settings and calibration status during subsequent measurements.

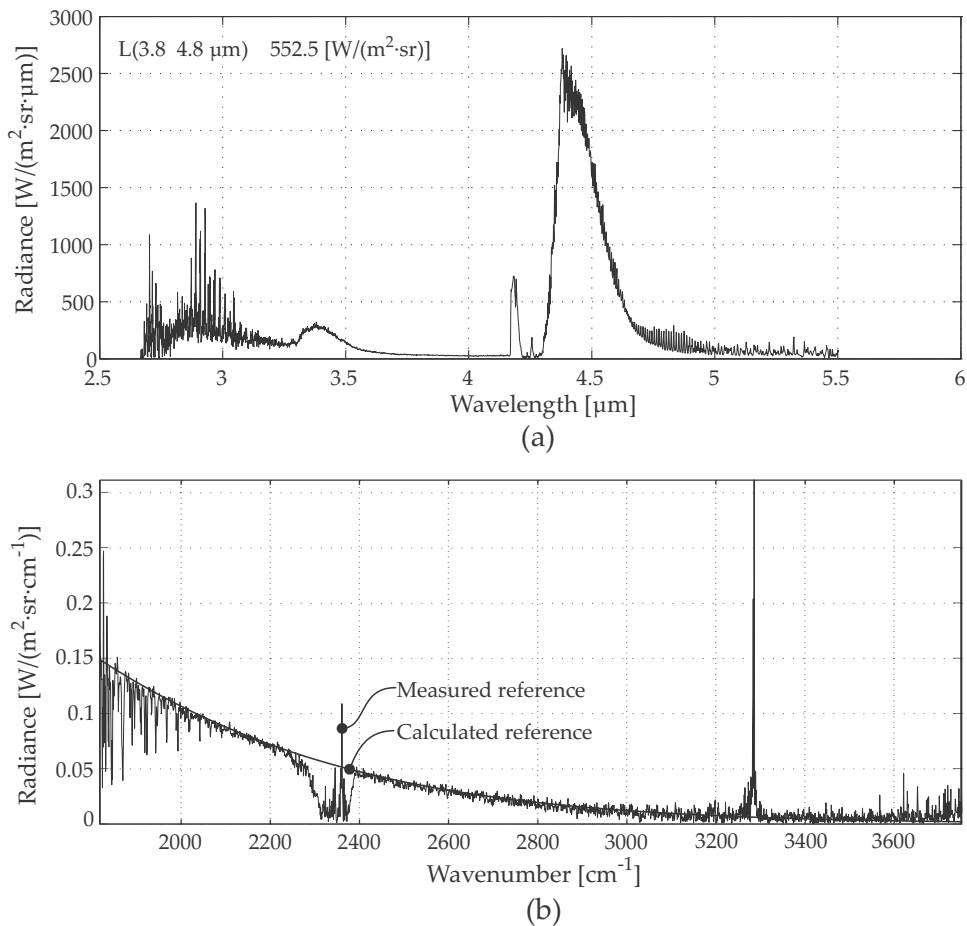


Figure 9.17 Spectral radiance for (a) the Bunsen burner yellow flame and (b) the reference source.

The FTIR instrument does not have the same fine spatial resolution as the imaging instruments, and the measurement represents some form of 'average' of the flame. The background temperature was much lower than the flame temperature and was ignored in further analysis. The FTIR spectrometer's small FOV was pointed at the bottom of the flame — a relatively cold part of the flame. Furthermore, the flame radiance was not uniform in the instrument's FOV. Because of the uncertainty in percentage fill and flame nonuniformity, there is some uncertainty in the absolute magnitude of the measurement. The radiance spectral shape was later shown to be accurate (see below). If the measured results are accepted as a scaled version of the full flame radiance, it is still useful because the shape of the spectral emissivity can be extracted. The top graph in Figure 9.17 shows the FTIR spectral radiance measurement as well as the wideband

integrated radiance in the MWIR band.

The bottom graph in Figure 9.17 shows the measured and calculated spectral radiance of a 150 °C blackbody reference source. Around 2350 cm^{-1} the atmospheric CO_2 absorption is very high, even over very short path lengths — as is visible in the graph. Around 3280 cm^{-1} an instrument anomaly is visible. It is evident that, apart from two anomalous bands, the two curves agree fairly well, thereby validating the instrument operation and hence the measurement.

9.8.4 Imaging-camera radiance results

The images are initially recorded as voltages or digital levels, proportional to the optical flux on the detector. The inverse of the calibration process (Section 9.8.2) provides radiance images, where a pixel represents the scene radiance (if an extended target fills the pixel). Using the spectral emissivity estimate, Equation (9.58), and the inverse form of Equation (9.59), the temperature for the pixel can be calculated. Inverting Equation (9.59) cannot be done analytically; a numerical solution must be found, such as a lookup table between radiance and temperature.

The results from this processing are shown in Figure 9.18. The top left graph shows the digital levels recorded by the instrument. The top right graph shows the object radiance, as calculated from the digital levels and the inverse calibration data. The images in Figure 9.18 demonstrate the gas flow and radiance/temperature in the burner flame. Near the bottom, the flame has a cool core of fresh gas supply with a higher radiance toward the edges, clearly indicating lower combustion activity in the center. Higher up, the air diffuses into the gas and provides oxygen to support combustion, depleting the amount of unburnt gas toward the top of the conical section.

As a confirmation check on the radiance levels, note that the total integrated radiance measured by the FTIR is 552.5 $\text{W}/\text{m}^2\text{sr}$ (Figure 9.17); this is the sum of all values measured by the FTIR in the MWIR band. The image data in Figure 9.18 (top right image) shows that the radiance levels immediately above the nozzle are of the order of 500–600 $\text{W}/\text{m}^2\text{sr}$, in good agreement with the FTIR radiance measurement. It appears that the concern for magnitude uncertainty in the FTIR measurement was unwarranted.

The two bottom graphs in Figure 9.18 should be read in conjunction: the left graph shows the temperature associated with the emissivity in the right graph. The critical assumption here is that the emissivity is the same

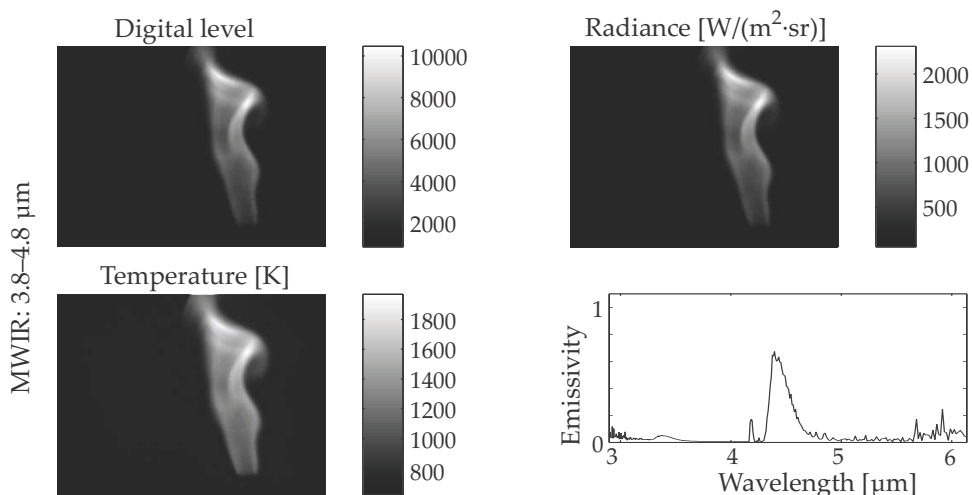


Figure 9.18 Bunsen flame MWIR images in units of digital level, radiance, and temperature.

at all temperatures; a more sophisticated analysis will allow for emissivity variation with temperature. It is also assumed that the emissivity is the same for all pixels in the image; this, too, requires a more-accurate model for advanced analysis.

Following the approach described in Section 8.4, various combinations of emissivity and temperature were considered. The top graph in Figure 9.19 shows the spectral emissivity and associated temperatures. Clearly, the case for a temperature of 650 °C is wrong because it requires an emissivity greater than one. Any of the remaining combinations are plausible but not necessarily physically feasible. If the flame temperature is 850 °C, the peak emissivity in the CO₂ spectrum is around 0.6. If the flame temperature is hotter, at 1250 °C the peak emissivity in the CO₂ spectrum is around 0.25. These combinations of temperature and emissivity all give the same radiance values as measured. After consulting published data,^{15–18} it was decided to select the 825 °C emissivity-temperature data set, shown in the bottom graph in Figure 9.19.

Section 4.2.5 describes the effect of optical thickness in a gaseous flame on the emissivity of the flame. Radiation measurements such as this can only give an indication of the apparent temperature as derived from the emissivity. The volumetric region with the highest emissivity will dominate the apparent temperature of the flame. An optically thick flame (high attenuation inside the flame) presents itself as a ‘surface’ radiator, and no conclusion can be reached on the temperature deep inside the flame. In this case, the flame is optically thin (emissivity less than unity), and the

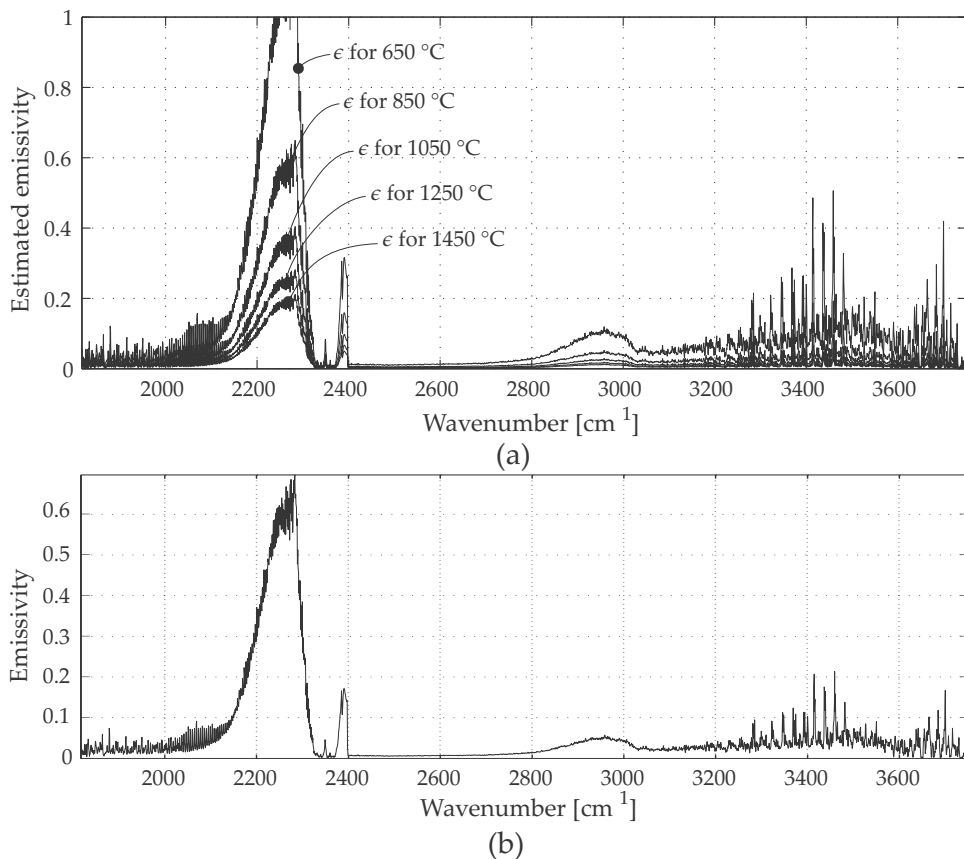


Figure 9.19 (a) Bunsen burner estimated emissivity-temperature combinations. (b) Apparent emissivity of the Bunsen burner yellow flame at 825 °C.

measurement is an indication of the temperature along the full path but dominated by the volumetric region with the highest emissivity.

At the bottom of the plume near the Bunsen burner nozzle, the temperature is around 1100 K (826 °C), whereas in the hottest region of the flame the temperatures are 1900 K (1626 °C). These values agree well with published information.^{15–18} The LWIR images were processed similarly to the MWIR images described above. The spectral emissivity was assumed to be constant over the LWIR band and found to be very low, less than 1%.

9.8.5 Imaging-camera flame-area results

The flame area is calculated from the radiance image using the techniques described in Section 8.4.2. The flame area was determined for nine radiance threshold values, ranging from just above background to 95% of the

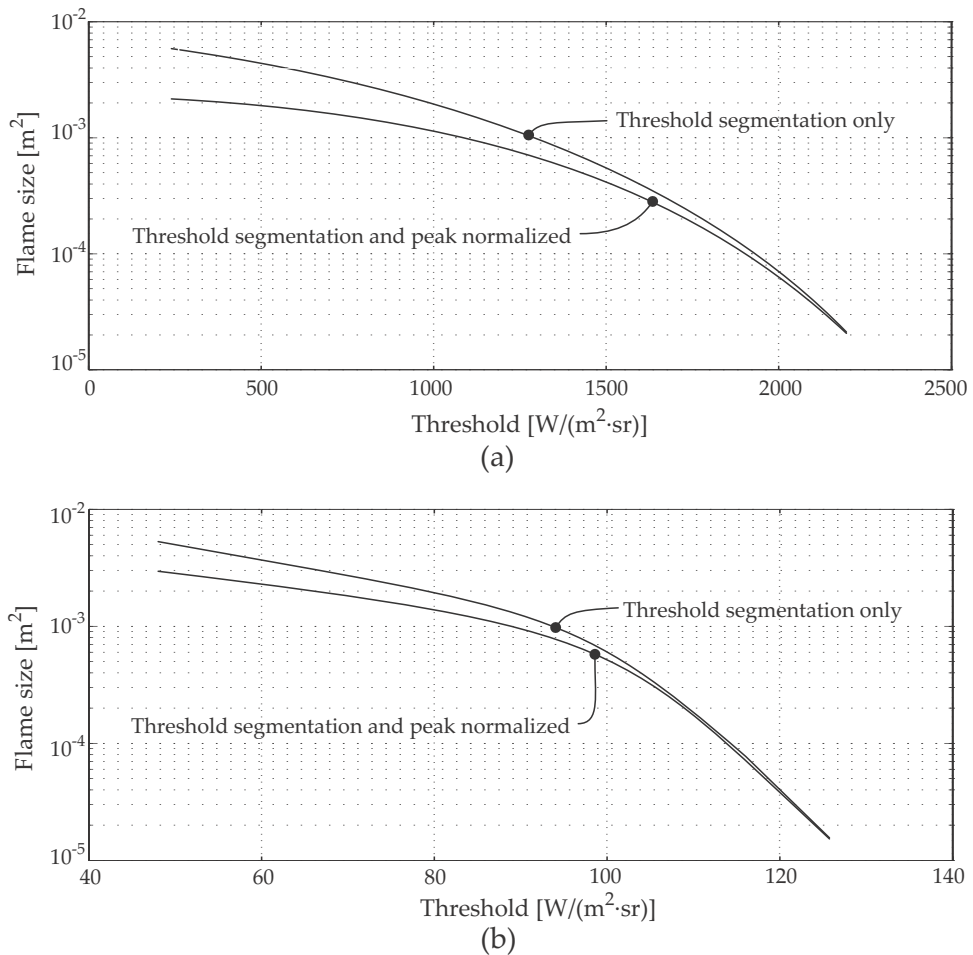


Figure 9.20 Bunsen burner yellow flame predicted area as function of threshold for (a) MWIR, and (b) LWIR.

maximum radiance. An analysis similar to the one depicted in Figure 8.5 was performed. The flame areas thus obtained from the MWIR and LWIR images are shown in Figure 9.20.

9.8.6 Flame dynamics

The Bunsen yellow flame has significant turbulence as the air mixes with the flame, consuming oxygen and expanding the hot gas plume. The forces resulting from internal mass flow acting on this plume cause it to become turbulent, resulting in the 'dancing' of the flame. The flame shape varied considerably in time as the turbulence contorted the flame. Figure 9.21 shows subsequent images in a measurement series; the interval between

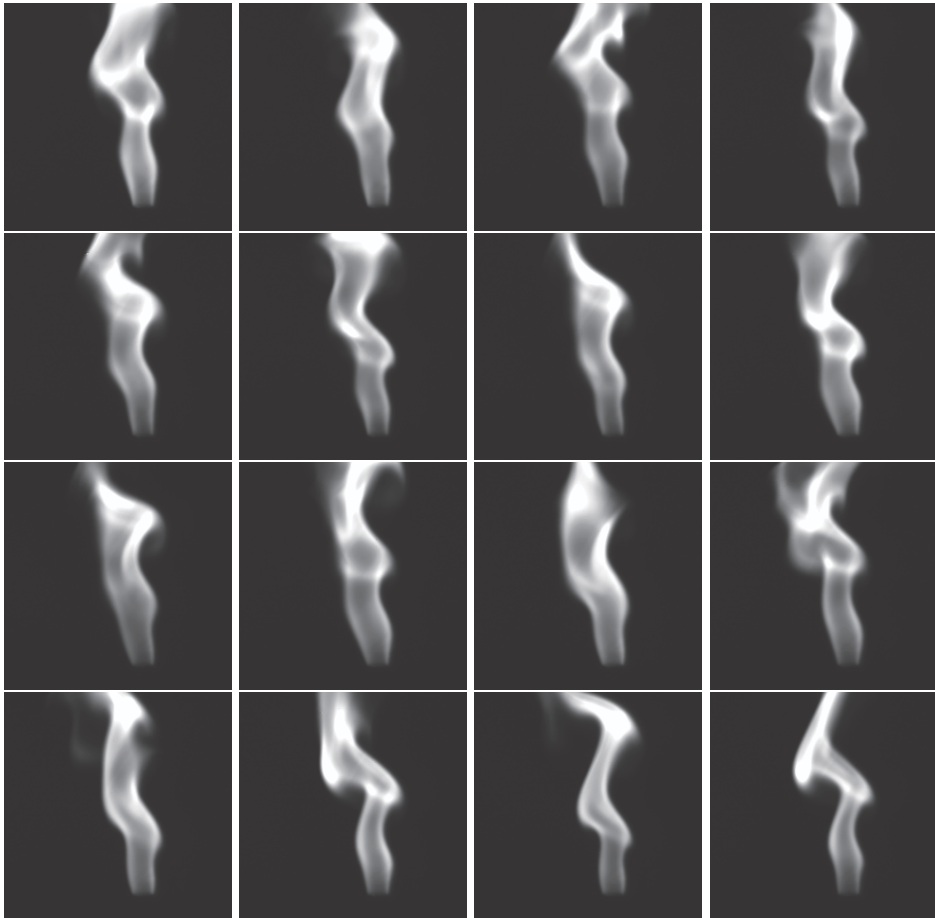


Figure 9.21 Bunsen flame sequence with 60-ms intervals between frames.

successive frames was 60 ms. The flame shapes vary considerably between subsequent frames. The turbulence bandwidth for this flame is higher than 20 Hz. There is little correlation between the ‘hot spots’ in subsequent frames. The hot spots indicate the spatial locations of combustion, and these locations change very quickly. The flame shapes show huge differences, but the flame volume (size) appears to be similar between all of the frames. The flame area is therefore expected to show only a small variation in time.

9.8.7 Thermocouple flame temperature results

The measurement of flame temperatures with a thermocouple¹⁹ proves to be quite difficult.^{20,21} Thermocouple readings are for the thermocouple itself, not necessarily for the flame. The indicated temperature varies greatly

from moment to moment. More importantly, the temperature reads lower than the expected adiabatic temperature due to ‘thermal loading’ — the thermocouple causes heat flow out of the (solid, liquid, or gas) test sample. Under thermal loading, the thermocouple never reaches equilibrium with the heat source, and the indicated temperature seems too low.

Thermocouple temperature measurements can be improved by reducing the thermal loading by using smaller thermocouple devices, using thinner wires, and by waiting for thermal equilibrium. In this experiment, thermal equilibrium was not achieved, and hence no conclusive result was obtained.

Bibliography

- [1] Pyradi team, “Pyradi Radiometry Python Toolkit,” <http://code.google.com/p/pyradi>.
- [2] Kaminski, W. R., “Range calculations for IR rangefinder and designators,” *Proc. SPIE* **227**, 65–79 (1980) [doi: 10.1117/12.958748].
- [3] RCA Corporation, *RCA Electro-Optics Handbook*, no. 11 in EOH, Burle (1974).
- [4] Daniels, A., *Infrared Systems, Detectors and FPAs*, 2 Ed., SPIE Press (2010).
- [5] Lloyd, J. M., *Thermal Imaging Systems*, Plenum Press, New York (1975).
- [6] Rogatto, W. D., Ed., *The Infrared and Electro-Optical Systems Handbook: Electro-Optical Components*, Vol. 3, ERIM and SPIE Press, Bellingham, WA (1993).
- [7] Rogalski, A., *Infrared Detectors*, 2nd Ed., CRC Press, Boca Raton, FL (2011).
- [8] Campana, S. B., Ed., *The Infrared and Electro-Optical Systems Handbook: Passive Electro-Optical Systems*, Vol. 5, ERIM and SPIE Press, Bellingham, WA (1993).
- [9] Wolfe, W. L. and Zissis, G., *The Infrared Handbook*, Office of Naval Research, US Navy, Infrared Information and Analysis Center, Environmental Research Institute of Michigan (1978).
- [10] Dereniak, E. L. and Boreman, G. D., *Infrared Detectors and Systems*, John Wiley & Sons, New York (1996).

- [11] Holst, G., *Electro-Optical Imaging System Performance*, 5th Ed., JCD Publishing, Winter Park, FL (2008).
- [12] Palmer, J. M. and Grant, B. G., *The Art of Radiometry*, SPIE Press, Bellingham, WA (2009) [doi: 10.1117/3.798237].
- [13] Slater, P., *Remote Sensing: Optics and Optical Systems*, Addison-Wesley, Boston, MA (1980).
- [14] Strojnik, M., Paez, G., and Granados, J. C., "Flame thermometry," *Proc. SPIE* **6307**, 63070L (2006) [doi: 10.1117/12.674938].
- [15] Haber, L. C., *An investigation into the origin, measurement and application of chemiluminescent light emissions from premixed flames*, Master's thesis, Virginia Polytechnic Institute and State University (2000).
- [16] Flame temperatures, <http://www.derosenet.net/steve/resources/engtables/flametemp.html>.
- [17] How Hot is a Bunsen Burner Flame?, http://www.avogadro-lab-supply.com/content.php?content_id=1003.
- [18] Wikipedia, "Flames," <http://en.wikipedia.org/wiki/Flame>.
- [19] Wikipedia, "Thermocouple," en.wikipedia.org/wiki/Thermocouple.
- [20] Yildirim, Z., *Self-defense of large aircraft*, Master's thesis, Naval Post-graduate School (2008).
- [21] Okamoto, N., "Overview of Temperature Measurement," http://www.engr.sjsu.edu/ndejong/ME_146.htm.
- [22] Pyradi team, "Pyradi data," <https://code.google.com/p/pyradi/source/browse>.

Problems

- 9.1 The purpose of this study is to verify the proposed FOV and to optimize the design to achieve maximum SNR. The data for this problem is given in the DP01.zip data file on the pyradi website.²²

A flame sensor has an aperture area of 0.005 m^2 and a proposed FOV of 10^{-5} sr . The InSb detector has a peak responsivity of 2.5 A/W and normalized spectral response defined in the data file (detectorNormalized), shown in Figure 9.22. The sensor filter

is defined in Equation (D.4) with $\tau_s = 0.0001$, $\tau_p = 0.9$, $\lambda_c = 4.3 \mu\text{m}$, $\Delta\lambda = 0.8 \mu\text{m}$, and $s = 12$.

The current flowing through the detector causes noise. The rms detector noise is given by $i_n = \sqrt{2qi_d\Delta f}$, where $q = 1.6 \times 10^{-19} \text{ C}$ is the charge on an electron, i_d is the DC current through the detector, and Δf is the noise equivalent bandwidth of the sensor electronics. For this problem, when calculating noise, ignore dark current through the detector; consider only the flux-induced current through the detector. The sensor's noise equivalent bandwidth Δf is 10^7 Hz .

The signal is defined as the difference between two samples: one sample with the flame (and sky) in the FOV, and one sample without the flame (only sky) in the FOV.

The system engineer requires a SNR of 8 to guarantee system performance.

The flame area is 0.1 m^2 , and the flame temperature is 1500°C . The flame emissivity is 0.1 over most of the spectral band due to carbon particles in the flame. At $4.3 \mu\text{m}$ there is a strong emissivity rise due to the hot CO_2 in the flame. The flame emissivity is defined in Equation (D.4) with the following parameters: $\tau_s = 0.1$, $\tau_p = 0.7$, $\lambda_c = 4.33$, $\Delta\lambda = 0.45$, and $s = 6$.

The distance between the flame and the sensor is 12 km. The atmospheric transmittance and sky radiance data files supplied with this problem are for the MODTRANTM Tropical climatic model. The path is oriented such that the sensor stares out to space at a zenith angle of 88 deg. The spectral transmittance is portrayed in Figure 9.22. The sky radiance file (skyRadiance) describes the sky/path radiance in units of $[\text{W}/(\text{cm}^2 \cdot \text{sr})]$, convert as appropriate. The transmittance file (tau12km) provides transmittance at the stated zenith angle for a range of 12 km.

- 9.1.1 Describe the possible means whereby the SNR can be optimized for this system. [3]
- 9.1.2 Use the data in the DP01.zip data file, or use MODTRANTM to calculate the transmittance and path radiance. Use the path geometry as defined above and confirm that the calculated values agree with the graphs shown here. [4]
- 9.1.3 Compile a mathematical formulation for the signal, the noise, and the SNR. This formulation will be used to evaluate your optimizing strategies, so it must be a complete description of the system with all parameters affecting system performance. [5]

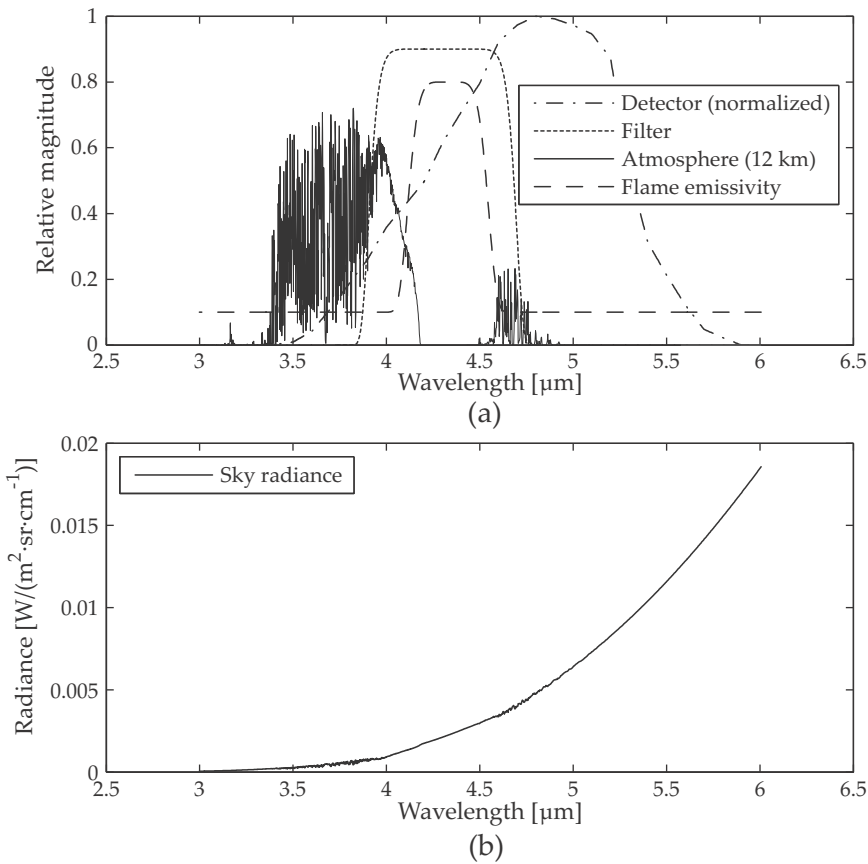
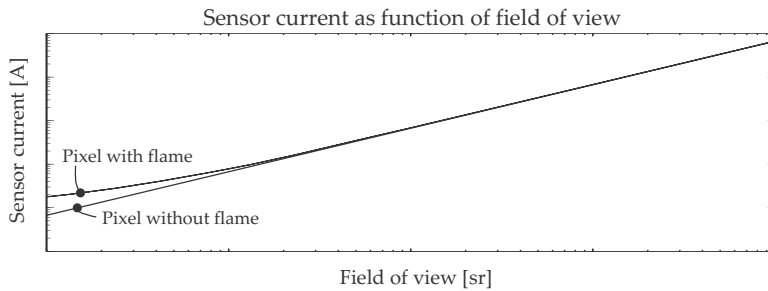


Figure 9.22 (a) Spectral data for the detector, filter, atmosphere, and flame emissivity. (b) Path radiance spectral data.

Apply the Golden Rules to the mathematical formulation derived here. [5]

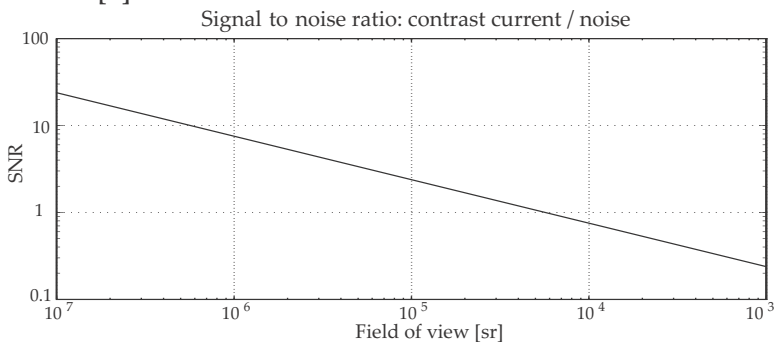
- 9.1.4 Write a numerical implementation of the mathematical model. Plot the spectral parameters and sky radiance, and verify that you get the same answers as in the graphs. [10]
- 9.1.5 Calculate the detector currents for sensor FOV values Ω of 1×10^{-7} , 1×10^{-6} , 1×10^{-5} , 1×10^{-4} , and 1×10^{-3} sr. [2]

Plot (1) the detector current with the flame in the FOV and (2) the detector current with the flame not in the FOV, versus FOV. Plot on log-log graphs; both the x and y axes must plot in log scale. Verify that you get curves of the form shown below. Explain the shape of the two curves. [4]



- 9.1.6 Calculate the SNR for the following sensor fields of view: 1×10^{-7} , 1×10^{-6} , 1×10^{-5} , 1×10^{-4} , and 1×10^{-3} sr. [2]

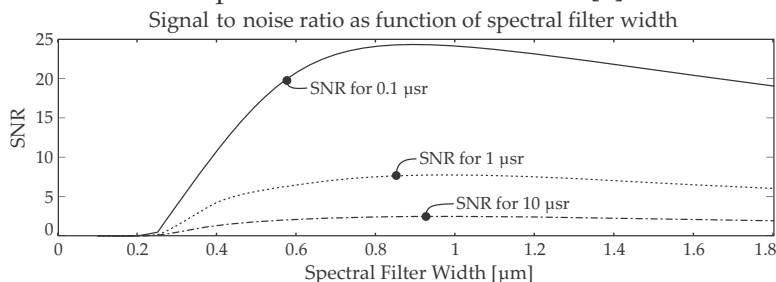
Plot the SNR versus FOV. Plot on log-log graphs and verify that you get a curve of the form shown below. Explain the shape of the curve. [2]



From this curve, determine whether a FOV of 1×10^{-5} sr will provide the necessary SNR, as specified by the system engineer. [1]

- 9.1.7 Calculate the SNR for each combination point on a two-dimensional grid of various spectral widths, starting at $0.1 \mu\text{m}$ and going to $0.8 \mu\text{m}$ in increments of $0.1 \mu\text{m}$ and fields of view of 1×10^{-7} , 1×10^{-6} , 1×10^{-5} , 1×10^{-4} , and 1×10^{-3} sr. Plot the SNR in a graph similar to the graph below. [4]

Assuming the graph below to be correct, comment on the optimal FOV and filter spectral width for the sensor. [2]



- 9.2 The data for this problem is given in the DP02.zip data file on the pyradi website.²²

A multi-color sensor operates in three bands: the visual band, the 3–5- μm band, and the 8–12- μm band. The sensor observes three infinitely large targets at a range of 2 km, with temperatures $T \in \{300, 1000, 6000\}$ K, with emissivity $\epsilon = 1$, through three different atmospheres. The purpose with this investigation is to study the target signatures in the different bands, and to calculate the detector signals in each case.

The atmospheric transmittance and path radiance are supplied as MODTRANTM files, in electronic format. The MODTRANTM results were calculated for a uniform path with length of 2 km and can be used as supplied. The MODTRANTM *tape5*, *tape6*, and *tape7* files contain the input data, user readable output data, and tabular data respectively. The files are as follows:

File	Description	Notes
atmo1	23-km visibility (Rural), 5 g/m ³ H ₂ O	Moderate conditions
atmo2	2-km visibility (Urban), 5 g/m ³ H ₂ O	Poor visibility
atmo3	23-km visibility (Rural), 39 g/m ³ H ₂ O	High humidity

In the *tape7* files, wavenumber is given in the 'FREQ' column, transmittance in the 'TOT TRANS' column, path radiance in the 'TOTAL RAD' column, and $-\gamma R$ is given in the 'DEPTH' column. Note that the path radiance is given in units of $[\text{W}/(\text{cm}^2 \cdot \text{sr} \cdot \text{cm}^{-1})]$.

The sensor has a FOV of 10^{-6} sr. The sensor FOV is completely filled by the target. The sensor optical aperture diameter is 100 mm. In this analysis you only have to consider the target flux and path radiance between the target and sensor — ignore background flux and flux inside the sensor.

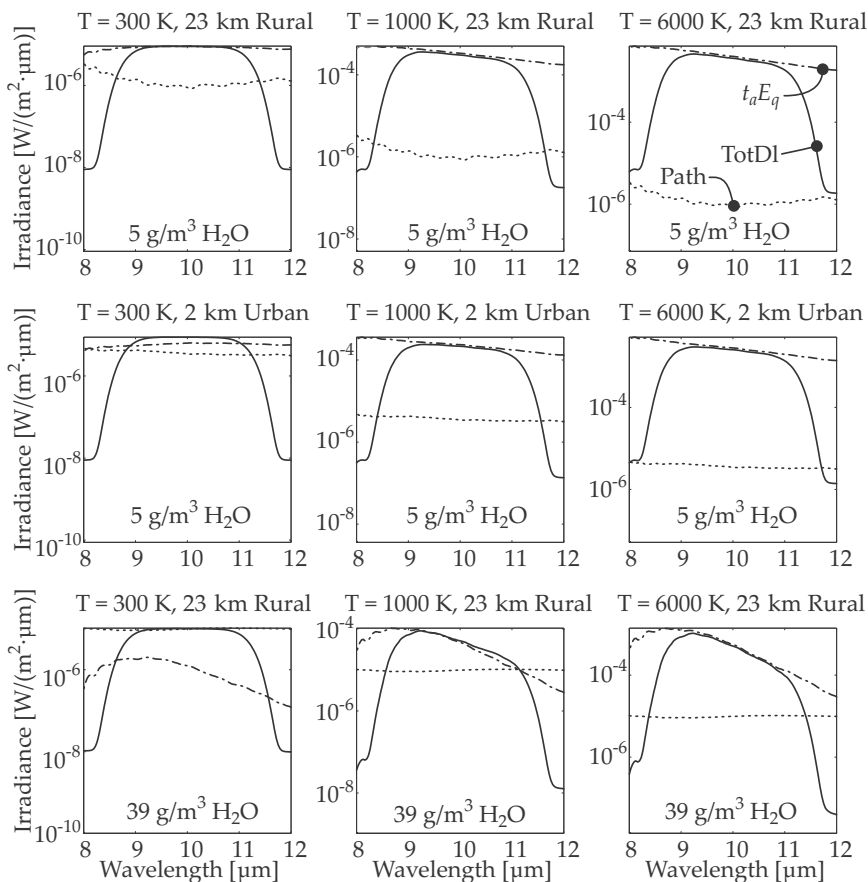
The detector spectral response is given by Equation (D.5) with the following values:

Band	λ_c	n	a	k
Silicon	1.20	4.30	3.50	8
3–6 μm	6.00	4.30	3.50	30
8–12 μm	12.00	4.30	3.50	60

An optical filter is used to limit the spectral width. The filter response is given by Equation (D.4) with the following values:

Band	τ_s	τ_p	s	$\Delta\lambda$	λ_c
Silicon	0.001	0.90	20	0.20	0.55
3–6 μm	0.001	0.90	12	1.20	4.20
8–12 μm	0.001	0.90	6	2.50	10.00

- 9.2.1 Draw a picture of the system. Write a mathematical formulation describing the detector current; include flux transfer, detector response, etc. Describe all elements in the model and provide the relevant numerical values for all parameters. [6]
- 9.2.2 Apply the Golden Rules to the mathematical formulation given above. [4]
- 9.2.3 Write a numerical implementation of the problem in a computer language. Describe the structure of the model and provide all numeric values (spectral and scalar). Use at least 100 samples in each spectral band at constant *wavenumber* increments. [6]
- Confirm the accuracy of your filter and detector implementations by plotting the spectral values. Confirm the correct read-in and processing of your atmospheric results by plotting the spectral atmospheric transmittance and path radiance. [4]
- 9.2.4 Compile a series of graphs of the spectral irradiance showing (1) the unfiltered path irradiance, (2) the targets' unfiltered thermal irradiance, and (3) the filtered sum of path plus target irradiance on the entrance aperture of the sensor (i.e., filter times the sum of the first two components). The graph must show the irradiance data vs. wavelength, with no detector weighting, for all combinations of sensors, atmospheres, and targets (27 graphs in total). The graphs should resemble those shown below but for all three spectral bands. 'Tot $\Delta\lambda$ ' is the total irradiance passed by the filter. 'Path' is the path radiance. ' $\tau_a E_\theta$ ' is the source thermal irradiance observed through the atmosphere. [6]



9.2.5 Calculate the current flowing through each of the detectors when viewing the different targets through the various atmospheres. Using the moderate atmosphere as a baseline, calculate the ratios of currents for the two adverse atmospheric conditions. Compare your results with the following table (current values are given in scientific notation and the ratio of currents are given in brackets). The results will not be exactly the same but should be of the same magnitude. [10]

Temperature = 300 K			
Atmosphere, H ₂ O content	Silicon	3–6 μm	8–12 μm
23-km visibility Rural, 5 g/m ³	1.09×10^{-25} (1.000)	1.87×10^{-8} (1.000)	6.47×10^{-7} (1.000)
2-km visibility Urban, 5 g/m ³	1.09×10^{-25} (1.000)	1.84×10^{-8} (0.982)	6.42×10^{-7} (0.992)
23-km visibility Rural, 39 g/m ³	1.09×10^{-25} (1.000)	2.20×10^{-8} (1.172)	7.28×10^{-7} (1.124)
	[A] (ratio)	[A] (ratio)	[A] (ratio)
Temperature = 1000 K			
Atmosphere, H ₂ O content	Silicon	3–6 μm	8–12 μm
23-km visibility Rural, 5 g/m ³	1.35×10^{-11} (1.000)	4.27×10^{-5} (1.000)	2.30×10^{-5} (1.000)
2-km visibility Urban, 5 g/m ³	5.97×10^{-13} (0.044)	2.63×10^{-5} (0.617)	1.60×10^{-5} (0.694)
23-km visibility Rural, 39 g/m ³	1.32×10^{-11} (0.972)	3.15×10^{-5} (0.737)	4.28×10^{-6} (0.186)
	[A] (ratio)	[A] (ratio)	[A] (ratio)
Temperature = 6000 K			
Atmosphere, H ₂ O content	Silicon	3–6 μm	8–12 μm
23-km visibility Rural, 5 g/m ³	7.18×10^{-3} (1.000)	1.73×10^{-3} (1.000)	2.79×10^{-4} (1.000)
2-km visibility Urban, 5 g/m ³	2.33×10^{-4} (0.033)	1.06×10^{-3} (0.614)	1.91×10^{-4} (0.685)
23-km visibility Rural, 39 g/m ³	7.09×10^{-3} (0.988)	1.29×10^{-3} (0.750)	4.57×10^{-5} (0.164)
	[A] (ratio)	[A] (ratio)	[A] (ratio)

- 9.2.6 Analyze the results obtained in the previous two questions. Comment on the relevance of each respective sensor to observe each target. Review the effect of the different atmospheric conditions on the observed irradiance values. Make recommendations as to which spectral bands to use for different sources and different atmospheric conditions. For example, complete a table such as shown below. [6]

Best spectral band for target and atmosphere			
	300 K	1000 K	6000 K
Moderate atmosphere	?	?	?
Low vis atmosphere	?	?	?
High humidity atmosphere	?	?	?

Finally, conclude on the effect of path radiance contribution to the total signature. [2]

- 9.3 Derive Equation (5.55) from Equation (5.58). [3]
- 9.4 Use Equation (5.55) to derive Equation (5.58). [3]
- 9.5 Use Equation (5.59) to derive Equation (5.60). [3]
- 9.6 Calculate the NETD of a thermal detector that is limited by temperature-fluctuation noise and photon noise only (no other noise sources). Both these two noise sources are present at the same time in the detector. Use Equation (5.55) and Equation (5.59) to derive an equation for the NEDT. Calculate and plot the NEDT versus G with the following detector parameters: a square pixel with dimensions $50\text{ }\mu\text{m} \times 50\text{ }\mu\text{m}$, absorption is 50%, pixel fill factor is 0.8, optics f -number is $f/1.5$, frame rate is 25 Hz. The detector does not have a cold shield. The background temperature is 300 K. Plot the results for a detector temperature of 77 K, 195 K and 300 K. [10]
- 9.7 Repeat problem 6 above, but add the effect of resistor Johnson noise. Evaluate the NETD for different values of the resistor value. [5]

Chapter 10

Golden Rules

The golden rule is that there are no golden rules.

George Bernard Shaw

10.1 Best Practices in Radiometric Calculation

Radiometric calculation can be rather tricky. The guidelines in this chapter are offered as a 'best practice' to help readers avoid preventable mistakes. Contrary to Mr. Shaw's statement, there may be a need for golden rules in radiometry!

10.2 Start from First Principles

Figure 10.1 summarizes almost everything one must remember when doing radiometric calculations.

Always view the problem as an application of Figure 2.11 and Equation (2.31); repeated as Figure 10.1. All problems can be rooted in this simple model: start here, and extend in wavelength, medium effects and/or geometry. In particular: (1) consider the source area as some surface in

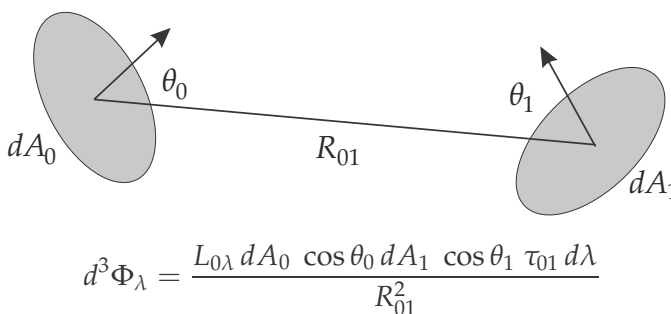


Figure 10.1 Flux transfer between two elemental surfaces.

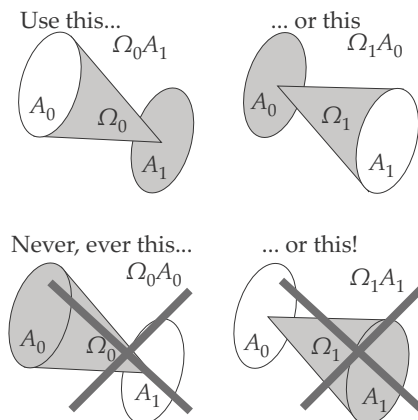


Figure 10.2 Legal combinations of solid angle and source area.

space and the receiver area as some surface in space, and integrate over both surfaces, (2) consider the spectral properties and integrate over wavelength or wavenumber, and (3) consider medium effects such as transmittance and path radiance.

Note that in the right of the flux transfer equation there are only two radiometric quantities L and λ and one medium property τ_{01} ; the remaining quantities are all geometric (nonradiometric) quantities. Radiometry is therefore as much a study of geometry as it is of optical flux. Get the geometry right, and the solution falls in place. If there is not a clear picture of the geometry, the correct solution is out of reach.

10.3 Understand Radiance, Area, and Solid Angle

It is easy to get confused by source area and receiver area — when to use which? The simple rule is to think of it, as if you are standing on one of the two surfaces and you are viewing the other surface. You cannot view the one you are standing on, you can only view the other. Your feet and eyes cannot rest on the same surface. This is shown Figure 10.2.

10.4 Build Mathematical Models

Derive a mathematical model from Equation (2.31). Start simple and add components as required by the problem at hand. Using the drawing as input, add factors for the atmosphere, lenses, optical filters, detectors, choppers, and amplifiers. Do not add factors for components that are not specified in the problem statement.

Table 10.1 SI base units.

Unit name	Unit symbol	Base quantity	Dimension symbol
meter	m	length	L
kilogram	kg	mass	M
second	s	time	T
ampere	A	electric current	I
kelvin	K	thermodynamic temperature	Θ
mole	mol	amount of substance	N
candela	cd	luminous intensity	J

10.5 Work in Base SI Units

As early as possible, convert problem units to base SI units,^{1,2} shown in Table 10.1, or directly-derived units. The value of a physical quantity can be expressed as the product of a numerical value (i.e., 4.3) and a unit (i.e., μm), both of which are algebraic factors, that can be manipulated by the rules of algebra. The symbol μm is related to the symbol m, by $\mu\text{m} = 1 \times 10^{-6} \text{ m}$, thus $\mu\text{m}/(1 \times 10^{-6}) = \text{m}$. One can therefore write $\lambda = 4.3 \mu\text{m}$, or divide both sides to obtain $\lambda/\mu\text{m} = 4.3$. Download and study the free IUPAC Green Book² for more information.

It happens frequently that distances and altitudes are given in [km] (e.g., the sun's diameter and distance in Section 3.7), temperature is in Celsius, or that detector D^* or sizes contains units of [cm]. Work in base SI units, rather than problem-domain units. Convert to SI units at the earliest possible time.

Some of the few exceptions to this rule are that wavelength is normally specified in [μm] or [nm], wavenumber is specified in [cm^{-1}], and D^* is specified in units of [$\text{cm} \cdot \sqrt{\text{Hz}/\text{W}}$].

10.6 Perform Dimensional Analysis

Test derived equations by manipulating or calculating the units for each of the variables. This is known as dimensional analysis^{2,3} or homogenous equation checking.⁴⁻⁷ This is a very effective method to ensure that you are using the correct areas in the flux calculation equations. It will also ensure that the equations and data are matched correctly.

Not all length dimensions [L] have the same meaning even though

they share the same SI unit: meter [m] (see Table 10.1 for dimensional symbols). Strict adherence to standards requires the use of the dimension symbols, such as [L] or [M], but using SI units [m] or [kg] works just as well. Subscript-mark all length dimensions with the meaning and location of such dimensions. For example, use $[m_0^2]$ or $[L_0^2]$ for source area, $[m_1^2]$ or $[L_1^2]$ for receiver area, $[m_d^2]$ or $[L_d^2]$ for detector area, and $[m_R]$ or $[L_R]$ for range. Use different subscripts for different surfaces, even if they all refer to area. Note that solid angle has units of $[m_0^2/m_R^2]$ or dimensions of $[L_0^2/L_R^2]$ when viewing the source (surface 0) from a sensor (surface 1). Once this is done, *do not* ‘cancel’ different types of lengths in the dimensional analysis, i.e., a m_s cannot ‘cancel’ a m_d because they are different types of length. Ensure that all appropriate units/dimensions are present and cancel correctly.

The dimensional analysis for Equation (2.31) is as follows:

$$d^2\Phi[W] = \frac{L dA_0 \cos\theta_0 dA_1 \cos\theta_1}{R^2} \left[\frac{W \cdot m_R^2}{m_0^2 \cdot m_1^2} \right] [m_0^2] [m_1^2] \left[\frac{1}{m_R^2} \right] \rightarrow [W],$$

or when considering spectral variables,

$$d^2\Phi[W] = \int_{\lambda} \frac{L_{\lambda} dA_0 \cos\theta_0 dA_1 \cos\theta_1 d\lambda}{R^2} \left[\frac{W \cdot m_R^2}{m_0^2 \cdot m_1^2 \cdot \mu m} \right] [m_0^2] [m_1^2] \left[\frac{1}{m_R^2} \right] \left[\frac{\mu m}{1} \right] \rightarrow [W].$$

Operations such as squares or square roots are also applied to units:

$$i_n = \sqrt{2qIB} \sqrt{\left[\frac{Q}{1} \right] \left[\frac{A}{1} \right] \left[\frac{1}{s} \right]} \rightarrow [A].$$

Be especially aware of constants with units, particularly the nonsymbol constants: they may not have symbols, but they certainly have units! Wavenumber is related to wavelength by $\lambda = 10^4/\tilde{\nu}$, where wavenumber has units of $[cm^{-1}]$, and wavelength has units of $[\mu m]$. The constant 10^4 here has units of $[cm^{-1} \cdot \mu m]$. The conversion of spectral densities between wavenumber and wavelength is given by

$$\frac{d\lambda}{d\tilde{\nu}} = -\frac{10^4}{\tilde{\nu}^2} [\mu m \cdot cm^{-1}] \left[\frac{1}{(cm^{-1})^2} \right] \rightarrow \left[\frac{\mu m}{cm^{-1}} \right].$$

10.7 Draw Pictures

Draw a picture of the system and the spatial geometry of the problem, such as shown in Figures 10.3 and 10.4. Write down what effect that component

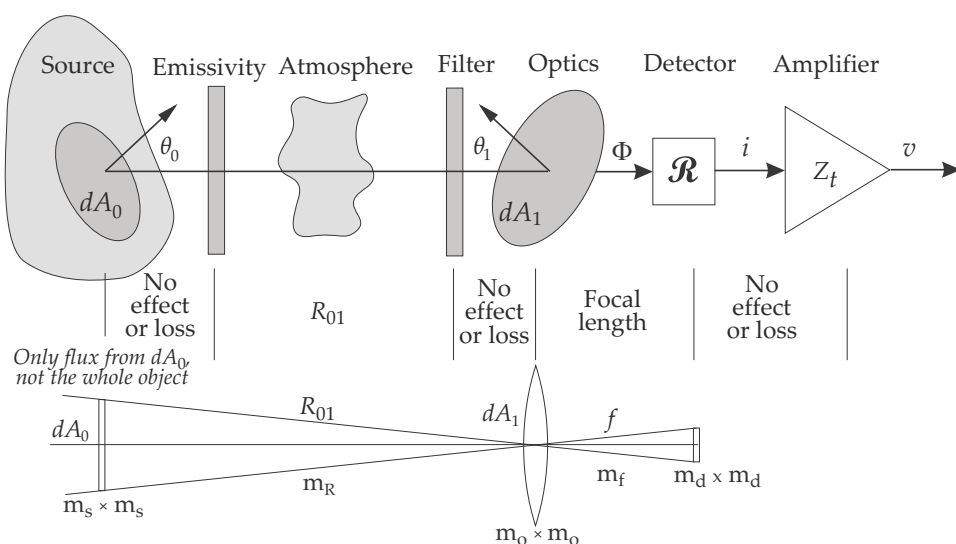


Figure 10.3 Diagram of a source, medium, and sensor.

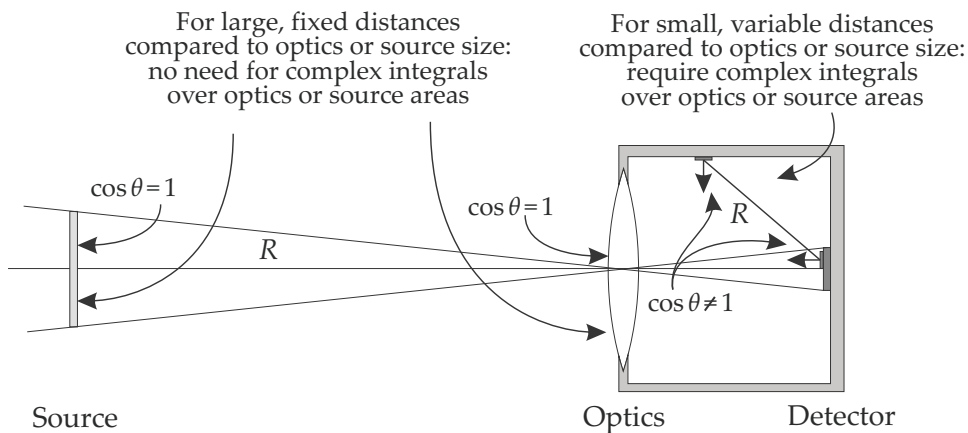


Figure 10.4 Spatial relationships in source and sensor.

Table 10.2 Component scaling and conversion.

Component	Value	Units	Converts
source area	1×10^{-5}	m_s^2	radiance \rightarrow intensity
emissivity	ϵ_λ	-	scales radiance spectrally
atmosphere	τ_λ	-	scales radiance spectrally
range	2500	m_R	R^2 : intensity \rightarrow irradiance
filter	τ_f	-	scales irradiance
optics area	1×10^{-6}	m_o^2	irradiance \rightarrow flux
detector	1.6	A/W	flux \rightarrow current
amplifier	2.5×10^4	V/A	current \rightarrow voltage
optical focal length	0.1	m_f	forms sensor field of view
detector size	0.001	m_d	forms sensor field of view
sensor FOV	1×10^{-4}	m_d^2/m_f^2	ω is the field of view
		m_s^2/m_R^2	ω : radiance \rightarrow irradiance

has on the flow or unit conversion of the flux or signal, and what effect the medium has on the flux/signal.

The drawing should specify which component provides the flux and which component receives the flux. It is very easy to confuse the optical aperture of a lens with the detector area as the receptor of flux.

Ensure that the drawing clearly shows not only the full extent or size of the source object but also which part is visible to the sensor. The sensor can only sense the flux radiating within its FOV — the object's radiation outside the sensor FOV does not contribute to the sensor signal.

For example, consider a system comprising a spectral source, an atmospheric medium, a lens, an optical filter, a detector, and an amplifier. The diagram could look like Figure 10.3.

In this picture, pay detailed attention to the source and receiving areas; in particular, make sure which part receives the flux. How is the flux transferred from one block to the next — are there losses or unit conversions along the way? The picture should also indicate if the cosine factors in Equation (2.31) can degenerate to unity or if these must be retained. Figure 10.4 illustrates these concepts.

For each of the objects in the figure, write down the type of component, value, the units, and the type of conversion taking place in the component. Consider the units of the component, i.e., [A/W] means the

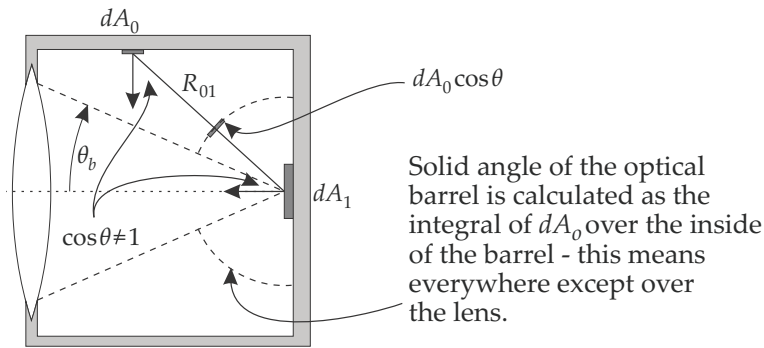


Figure 10.5 Spatial integration of source surface area.

component receives watts and outputs amperes. An area with units $[m^2]$ can convert irradiance in $[W/m^2]$ to flux $[W]$. An example is shown in Table 10.2.

It may also be beneficial to draw the shapes of the source and receiver true to the real-world object (e.g., the barrel containing the optical elements and the detector). Such a drawing will immediately indicate if it is necessary to perform a spatial integral $A = \int_A dA$ over one or both of the areas in Equation (2.31).

10.8 Understand the Role of π

Remember when to remember π . See Section 2.7 for the relationship between exitance and radiance for a Lambertian radiator. When working from first principles, there is no need to remember when to use π because it is taken care of in the mathematics.

The scripting Planck-law functions given in Appendix D provide exitance in $[W/m^2]$, not radiance in $[W/(m^2 \cdot sr)]$. When using these functions, divide the result by π to get radiance.

10.9 Simplify Spatial Integrals

Simplify spatial integrals where possible. In Figure 10.5 the solid angle of the optical barrel is integrated over the inside of the box except over the lens. On the assumption that the optical barrel has uniform radiance, the barrel geometry can be 'collapsed' onto, and integrated over, the portion of the sphere shown in the figure. The spherical portion is rotationally symmetrical around the optical axis. In this case the projected solid angle of the barrel is given by $\omega = \pi - \pi \sin^2 \theta_b$.

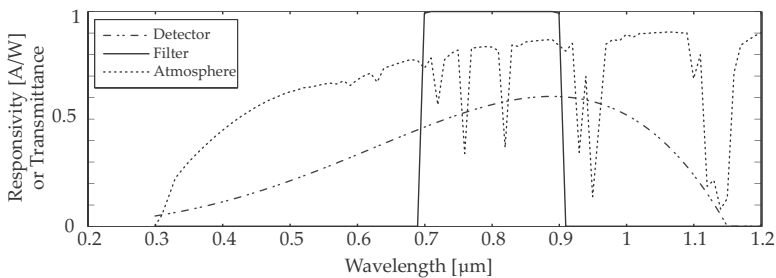


Figure 10.6 Graphical depiction of spectral variables.

10.10 Graphically Plot Intermediate Results

To confirm visually that the calculation is correct, plot the calculated values graphically and inspect the graphs very carefully. For example, Figure 10.6 displays filter and detector spectral responses that were calculated and the atmospheric transmittance loaded from file; all are plotted to confirm that no error was made.

10.11 Follow Proper Coding Practices

When you code your problem in a computer language, ensure that you copy and update the code accurately. It happens too often that code is copied but the variables are not updated in the new context. The better solution is to use functions for repeating calculations — the benefit is that a change made once will apply to all use cases, and varying data can be clearly defined as function parameters.

Keep track of the value of constants' exponents and the e-notation in the computer scientific format: $10\text{e}4$ is $10 \times 10^4 = 10^5$ and not $10^4 = 1 \times 10^4$.

10.12 Verify and Validate

Even the most rudimentary radiometry calculation can go wrong. Find ways to verify and validate the results. Section B.3 describes a modeling verification and validation framework. In your calculation, identify the conceptual model and computer (calculation) model and carefully qualify, verify, and validate the processes and results along the calculation chain. Several techniques are available^{8,9} for this purpose.

10.13 Do It Right — the First Time!

Refrain from taking shortcuts or out-of-the-expected actions when doing an ‘initial’ investigation or calculation. These actions tend to be costly in the long run. When executing a task, spend just a little more time to safeguard your current action against future events. In many cases, you would have to revisit the work later, or you might want to re-apply today’s work products in another task. Do it right the first time!

In the context of this book, ‘doing it right — the first time’ means developing a proper model from first principles instead of using a quickie formula — quite possibly incorrectly! It means writing code instead of repeatedly typing the same thing in a command window. It means creating concepts, design packages, and computer code functions for re-use instead of endless copy-and-paste repeats. It means developing a toolset today for use tomorrow and the day after. Most importantly, it means documenting and archiving your work, adding it to a living and growing repository. All of this is done to obtain good return on your investment of time and effort.

Some reasons why a little more effort may be warranted: (1) There may be a risk in fixing a task in future; better fix it while doing it the first time. (2) There are additional costs in fixing a task later; you have to repeat the start-up context-switching activities. (3) Do the ‘expected’ — execute, document, and archive in a manner, place, or method that most people would expect — because it will ease the work for others, and even for yourself, later. (4) Keep in mind that you are most likely the future client of your current work, so be kind to yourself!

Bibliography

- [1] ISO, “Quantities and units – Part 1: General,” Standard ISO 80000-1:2009, International Organization for Standardization (2009).
- [2] Cohen, E., Cvitas, T., Frey, J., Holmström, B., Kuchitsu, K., Marquardt, R., Mills, I., Pavese, F., Quack, M., Stohner, J., Strauss, H., Takami, M., and Thor, A., *Quantities, Units and Symbols in Physical Chemistry*, 3rd Ed., IUPAC Green Book, IUPAC & RSC Publishing, Cambridge, UK (2008).
- [3] Wikipedia, “Dimensional analysis,” en.wikipedia.org/wiki/Dimensional_analysis.
- [4] Wikipedia, “SI Units,” en.wikipedia.org/wiki/SI.
- [5] NIST, “SI base units,” physics.nist.gov/cuu/Units/units.html.

-
- [6] Wikibooks, "The SI System of Units," en.wikibooks.org/wiki/A-level_Physics/The_SI_System_of_Units.
 - [7] Wright, D., "Units, dimensions and conversion factors," www.mech.uwa.edu.au/DANotes/units/units.html.
 - [8] Schlesinger, S., "SCS Technical Committee on Model Credibility: Terminology for Model Credibility," *Simulation* 32, 103–104 (1979).
 - [9] Sargent, R. G., "Verification and Validation of Simulation Models," *Proc 1998 Winter Simulation Conference* 1, 121–130 (1998).



Cornelius J. (Nelis) Willers completed a B.Eng (Honns) Electronics Engineering degree at the University of Pretoria in 1976 and an MS (Optical Engineering) degree at the University of Arizona in 1983. He is registered as a professional engineer. His 36 years of work experience includes electro-optical system development, system architecture and systems engineering, software development, and infrared scene simulation. His most notable achievements include be-

ing the chief architect and technical lead in establishing an imaging missile seeker technology base, and in the process, spearheading advanced physics-based infrared image simulation. The simulation system is currently used for a number of different applications in laboratories across the globe. His current interests include infrared signature measurement and data analysis, infrared system modeling and simulation, and the development of aircraft self-protection systems. He is leading the open-source, Python-based *pyradi* radiometry toolkit project. He has published a large number of technical and research reports. His conference paper topics include infrared system modeling and simulation, and the modeling of military conflict using agent-based techniques. He teaches radiometry and infrared system design in short courses and at a masters-degree level at the University of Pretoria.

Nomenclature

α	Absorptance, absorptivity, absorption (fraction)
α	Absorption attenuation coefficient with units [m^{-1}]
α_λ	Spectral absorption with units [m^{-1}]
α_B	Temperature coefficient of resistance with units [K^{-1}]
β	Diode p-n junction nonideal factor (unitless)
β	Optical thickness (unitless)
γ	Attenuation coefficient with units [m^{-1}]
Γ	Γ point: smallest energy difference in bandgap (condition)
$\delta()$	Dirac delta function (unitless)
Δ_ϵ	Spatial texture variation in emissivity (unitless)
Δ_ρ	Spatial texture variation in reflectivity (unitless)
$\Delta\Phi$	Change in optical flux with units [W] or [q/s]
$\Delta\Phi_e$	Change in radiant optical flux with units [W]
$\Delta\Phi_p$	Change in optical photon flux with units [q/s]
Δf	Noise equivalent bandwidth with units [Hz]
Δn_e	Change in number of electrons with units [quanta]
Δn_h	Change in number of holes with units [quanta]
ΔT	Change in temperature with units [K]
ϵ	Emissivity (unitless)
ϵ	Electric field across a distance with units [V/m]
ϵ_λ	Spectral emissivity (unitless)
η	Detector quantum efficiency (unitless)
η_a, η_b	Image fill efficiency along the a and b directions (unitless)
η_s	Scanning efficiency in an image-forming system (unitless)
θ	Angle with units [rad]
θ	Dimensional symbol for temperature, or thermal (unitless)
λ	Wavelength with units [μm]
λ_c	Cutoff wavelength with units [μm]
μ	Carrier mobility with units [$\text{cm}^2/(\text{s}\cdot\text{V})$]
μ_e	Electron carrier mobility with units [$\text{cm}^2/(\text{s}\cdot\text{V})$]
μ_h	Hole carrier mobility with units [$\text{cm}^2/(\text{s}\cdot\text{V})$]
ν	Frequency with units [Hz] or [s^{-1}]
$\tilde{\nu}$	Wavenumber with units [cm^{-1}]
ρ	Material density with units [g/m^3]

ρ	Reflectance, reflectivity, reflection (fraction)
ρ_λ	Spectral reflection (unitless)
ρ_d	Diffuse reflection (unitless)
ρ_s	Specular reflection (unitless)
σ	Material electrical conductivity with units [Ω/m]
σ	Scattering attenuation coefficient with units [m^{-1}]
σ	Surface roughness (root-mean-square) with units [m]
σ_e	Stefan–Boltzmann constant with units [$\text{W}/(\text{m}^2 \cdot \text{K}^4)$]
σ_q	Stefan–Boltzmann constant with units [$\text{q}/(\text{s} \cdot \text{m}^2 \cdot \text{K}^3)$]
τ	Transmittance, transmissivity, transmission (fraction)
τ_λ	Spectral transmittance (unitless)
τ_θ	Thermal time constant with units [s]
τ_a	Atmospheric transmittance (unitless)
τ_c	Contrast transmittance (unitless)
τ_e	Electron lifetime with units [s]
τ_h	Hole lifetime with units [s]
τ_{RC}	Electronic resistor–capacitor time constant with units [s]
Φ	Optical flux with units [W] or [q/s]
Φ_λ	Optical flux spectral density with units [$\text{W}/\mu\text{m}$]
Φ_e	Radiant optical flux with units [W]
Φ_p	Optical photon flux with units [q/s]
Φ_q	Optical photon flux with units [q/s]
ψ	Solar irradiance geometry factor with units [sr/sr]
ψ	Wave function for a free electron (unitless)
ω	Electrical frequency with units [rad/s]
ω	Geometric solid angle with units [sr]
ω	Pixel field of view solid angle with units [sr]
Ω	Projected solid angle with units [sr]
Ω_r	Field of regard in an image-forming system with units [sr]
A	Area with units [m^2]
A_d	Detector area with units [m^2]
A_s	Source area in units [m^2]
A_v	Voltage gain of an amplifier or filter with units [V/V]
BRDF	Bidirectional reflection distribution function with units [sr^{-1}]
c	Specific heat with units [$\text{J}/(\text{g} \cdot \text{K})$]
c	Speed of light in vacuum with units [m/s]
C	Contrast (unitless)
C, C_s	Thermal detector element heat capacity with units [J/K]
C_v	Contrast threshold (unitless)
CODATA	Committee on Data for Science and Technology
D	Diameter of an optical aperture or lens with units [m]
D	Detectivity with units [W^{-1}]

D	Diffusion constant with units [m^2/s]
D^*	Specific detectivity with units [$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$]
D_λ^*	Spectral specific detectivity with units [$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$]
D_{eff}^*	Wideband specific detectivity with units [$\text{cm} \cdot \sqrt{\text{Hz}}/\text{W}$]
D_e	Diffusion constant for electrons with units [cm^2/s]
D_h	Diffusion constant for holes with units [cm^2/s]
e	Electron with charge q with units [C]
E	Energy (semiconductor energy level) with units [J] or [eV]
E	Irradiance (Areance) with units [W/m^2]
E_λ	Irradiance (Areance) spectral density with units [$\text{W}/(\text{m}^2 \cdot \mu\text{m})$]
E_C	Lowest conduction band energy level with units [J] or [eV]
E_F	Fermi level with units [J] or [eV]
E_g	Semiconductor energy bandgap with units [J] or [eV]
E_q	Background photon flux with units [$q/(\text{s} \cdot \text{m}^2)$]
E_V	Highest valence band energy level with units [J] or [eV]
f	Electrical frequency with units of [Hz]
f	Focal length with units [m]
F	View factor or configuration factor with units [sr/sr]
f_{fill}	Fill factor, fraction of area filled (unitless)
F_F	Frame rate in an image-forming system with units [Hz]
f_r	Bidirectional reflection distribution function with units [sr^{-1}]
F_T	Fourier transform
$f_{-3 \text{ dB}}$	−3 dB electronic bandwidth with units [Hz]
$f/\#$	F -number, alternative notation (unitless)
$F_\#$	F -number of a lens, with numerical value $\#$ (unitless)
FAR	False alarm rate with units [s^{-1}]
FOM	Figure of merit
FOV	Field of view with units [rad]
FTIR	Fourier transform infrared
G	Detector photon gain with units [electrons/photon]
G	Heat conductance with units [W/K]
G_c	Bias circuit gain (unitless)
G_{ph}	Photoconductive gain with units [electrons/photon]
g_{th}	Rate of thermal carrier generation with units [quanta/s]
h	Planck constant with units [J·s]
\hbar	$\hbar = h/(2\pi)$ with units [J·s], where h is the Planck constant
i	Current with units [A]
i	Noise current density with units [$\text{A}/\sqrt{\text{Hz}}$]
I	Intensity (Pointance) with units [W/sr]
\hat{I}	Incident ray unit vector (unitless)
I_0	Reverse-bias-saturation current with units [A]
I_λ	Intensity (Pointance) spectral density with units [$\text{W}/(\text{sr} \cdot \mu\text{m})$]

I_b	Bias current with units [A]
i_{gr}	Generation–recombination noise with units [A] or $[A/\sqrt{\text{Hz}}]$
i_n	Noise current with units [A] or $[A/\sqrt{\text{Hz}}]$
I_{ph}	Photocurrent with units [A]
I_{sat}	Reverse-bias-saturation current with units [A]
J	Diffusion current density with units $[A/m^2]$
J_d	Drift current density with units $[A/m^2]$
k	Boltzmann constant with units [J/K]
K_λ	Spectral photopic luminous efficacy with units [lm/W]
K'_λ	Spectral scotopic luminous efficacy with units [lm/W]
K_μ	Sky-ground radiance ratio in thermal spectral bands (unitless)
K_v	Sky-ground radiance ratio in the visual spectral band (unitless)
k_f	Time-bandwidth product with units [s·Hz]
k_F	Reciprocal lattice sphere radius with units [m]
k_n	Ratio of noise equivalent bandwidth to -3 dB bandwidth
L	Radiance (Sterance) with units $[W/(m^2 \cdot sr)]$
L_λ	Radiance (Sterance) spectral density with units $[W/(m^2 \cdot sr \cdot \mu m)]$
L_v	Diffusion length for carriers with units [cm]
L_e	Diffusion length for electrons with units [cm]
L_h	Diffusion length for holes with units [cm]
L_p	Detector packaging inductance with units [H]
LWIR	Long-wave infrared
m	Mass with units [g] or [kg]
M	Exitance (Areance) with units $[W/m^2]$
M_λ	Exitance (Areance) spectral density with units $[W/(m^2 \cdot \mu m)]$
M_e	Radiant exitance with units $[W/m^2]$
m_e	Electron mass with units [g]
m_e^*	Effective electron mass in units of m_e
m_h^*	Effective hole mass in units of m_e
MDT	Minimum detectable temperature with units [K]
MRT	Minimum resolvable temperature with units [K]
MTF	Modulation transfer function
MTV	Magnesium-Teflon [®] -Viton [®]
MWIR	Medium-wave infrared
n	Electron concentration with units $[cm^{-3}]$
n	Index of refraction (unitless)
N	Number of objects, pixels, or detector elements (unitless)
\hat{N}	Surface normal unit vector (unitless)
n_a	Acceptor concentration with units $[cm^{-3}]$
n_d	Donor concentration with units $[cm^{-3}]$
n_e	Number of electrons (unitless)
n_h	Number of holes (unitless)

n_i	Intrinsic carrier concentration with units [cm^{-3}]
n_n	Electron concentration in n-type material with units [cm^{-3}]
n_p	Electron concentration in p-type material with units [cm^{-3}]
n_r	Real component of the complex index of refraction (unitless)
NA	Numerical aperture (unitless)
$\text{NE}\Delta\rho$	Noise equivalent reflectance (unitless)
$\text{NE}\Delta T$	Noise equivalent temperature difference with units [K]
NEE	Noise equivalent irradiance with units [W/m^2]
NEL	Noise equivalent radiance with units [$\text{W}/(\text{m}^2\cdot\text{sr})$]
NEM	Noise equivalent exitance with units [W/m^2]
NEP	Noise equivalent power with units [W]
NER	Noise equivalent reflectance (unitless)
NETC	Noise equivalent target contrast with units [K]
NETD	Noise equivalent temperature difference with units [K]
NIR	Near infrared
OTF	Optical transfer function
p	Hole concentration with units [cm^{-3}]
$P(\theta)$	Scattering phase function (unitless)
P_d	Probability of detection (unitless)
p_n	Hole concentration in n-type material with units [cm^{-3}]
P_n	Probability of false detection (unitless)
p_p	Hole concentration in p-type material with units [cm^{-3}]
PSD	Power spectral density with units [A^2/Hz] or [V^2/Hz]
PSF	Point spread function (unitless)
q	Absolute humidity with units [g/m^3]
q	Electron charge with units [C]
q	Quanta, as in photon count (unitless)
Q	Energy with units [$\text{W}\cdot\text{s}$] or [J]
r	Radius with units [m]
R	Range or distance with units [m]
\mathcal{R}	Responsivity with units [A/W] or [V/W]
$\hat{\mathcal{R}}$	Mirror reflection unit vector (unitless)
$\hat{\mathcal{R}}$	Detector responsivity scaling factor with units [A/W] or [V/W]
$\bar{\mathcal{R}}$	Equivalent path length with units [m]
$\tilde{\mathcal{R}}$	Normalized spectral shape of spectral responsivity (unitless)
\mathcal{R}_λ	Detector spectral responsivity with units [A/W] or [V/W]
R_0	Dynamic resistance under zero-bias conditions with units [Ω]
R_d	Detector resistance with units [Ω]
$\mathcal{R}_{e\lambda}$	Spectral detector responsivity with units [A/W] or [V/W]
$\mathcal{R}_{q\lambda}$	Spectral detector responsivity with units [C] or [J/A]
R_{eff}	Effective (wideband) responsivity with units [A/W] or [V/W]
R_L	Load resistor or bias resistor with units [Ω]

R_V	Meteorological range (visibility) with units [km]
RH	Relative humidity, unitless expressed as %
rms	Root-mean-square (unitless)
S, S_1, S_2	Seebeck coefficients for thermoelectricity with units [V/K]
\mathcal{S}	Sensor response
\hat{S}	Reflected ray unit vector (unitless)
S_λ	Sensor spectral response (unitless)
$S(\omega), S(f)$	Power spectral density with units [A^2/Hz] or [V^2/Hz]
SCR	Signal-to-clutter ratio (unitless)
SNR	Signal-to-noise ratio (unitless)
SWIR	Short-wave infrared
t	Time with units [s]
T	Temperature with units [K]
T	Throughput or étendue with units [$sr \cdot m^2$]
T_b	Background temperature [K]
T_{filter}	Temperature of an optical filter with units [K]
t_p	Signal pulse width with units [s]
T_s	Source temperature with units [K]
TPM	Technical performance measure
v	Voltage (signal or noise) with units [V]
V	Volume with units [m^3]
V_λ	Spectral photopic luminous efficiency (unitless)
V'_λ	Spectral scotopic luminous efficiency (unitless)
V_{bias}	Bias voltage across a device with units [V]
V_d	Internal potential in a p-n diode with units [V]
v_n	Noise expressed as voltage with units [V] or [V/\sqrt{Hz}]
w	Energy density with units [J/m^3]

Appendix A

Reference Information

Table A.1 Definition of radiometric quantities.

Quantity	Direction	Symbol	Definition	Basic Unit
Energy		Q	-	[W·s]
Flux		Φ	$\Phi = \frac{dQ}{dt}$	[W]
Density		w	$w = \frac{dQ}{dV}$	[W·s/m ³]
Intensity (Pointance)	exitent	I	$I = \frac{d\Phi}{d\omega} = \frac{d^2Q}{dt d\omega}$	[W/sr]
Exitance (Areance)	exitent	M	$M = \frac{d\Phi}{dA}$	[W/m ²]
Irradiance (Areance)	incident	E	$E = \frac{d\Phi}{dA}$	[W/m ²]
Radiance (Sterance)	spatial	L	$L = \frac{d^2\Phi}{d\omega \cos \theta dA}$	[W/(m ² ·sr)]
where:				
exitent		Energy leaving the surface		
incident		Energy falling in on a surface		
spatial		Energy anywhere in space		
time		t	seconds	[s]
volume		V	volume	[m ³]
solid angle		ω	steradian	[sr]
area		A	area	[m ²]
difference operator		d, d^2		
Basic Unit is defined in any of:			radiant	[W]
			photon rate	[q/s]
			photometric	[lm]

Table A.2 Physical and mathematical constants.^{1,2}

	Name	Magnitude	Units
Physical constants			
h	Planck's constant	$6.62606957 \times 10^{-34}$	J·s
c	Speed of light in vacuum	2.99792458×10^8	m/s
k	Boltzmann's constant	$1.3806488 \times 10^{-23}$	J/K
q	Electron charge	$1.602176565 \times 10^{-19}$	C
	Absolute zero (0 K)	−273.15	°C
	Melting point of ice	≈273.15	K
	Triple point of water	273.16	K
Mathematical constants			
π	pi	3.14159265358979	
e	Base of the natural logarithm	2.71828182845905	
$\zeta(3)$	Apéry's constant	1.2020569031595942853	
a_2	Solution to $2(1 - e^{-x}) - x = 0$	1.59362426004004	
a_3	Solution to $3(1 - e^{-x}) - x = 0$	2.82143937212208	
a_4	Solution to $4(1 - e^{-x}) - x = 0$	3.92069039487289	
a_5	Solution to $5(1 - e^{-x}) - x = 0$	4.96511423174429	

Table A.3 Planck-law constants.³

	Name	Magnitude	Units
Radiation constants			
c_{1e}	First radiation constant ($2\pi hc^2$)	$3.74177152466413 \times 10^{-16}$	$\text{W} \cdot \text{m}^2$
c_{1q}	First radiation constant ($2\pi c$)	$1.88365156730885 \times 10^9$	$\text{q} \cdot \text{m}/\text{s}$
c_2	Second radiation constant (hc/k)	$1.43877695998382 \times 10^{-2}$	$\text{m} \cdot \text{K}$
Radiation constants, λ expressed in spectral units of $[\mu\text{m}]$			
$c_{1e\lambda}$	First radiation constant ($2\pi hc^2$)	$3.74177152466413 \times 10^8$	$\text{W} \cdot \mu\text{m}^4/\text{m}^2$
$c_{1q\lambda}$	First radiation constant ($2\pi c$)	$1.88365156730885 \times 10^{27}$	$\text{q} \cdot \mu\text{m}^3/(\text{s} \cdot \text{m}^2)$
$c_{2\lambda}$	Second radiation constant (hc/k)	$1.43877695998382 \times 10^4$	$\mu\text{m} \cdot \text{K}$
Radiation constants, $\tilde{\nu}$ expressed in spectral units of $[\text{cm}^{-1}]$			
$c_{1e\tilde{\nu}}$	First radiation constant ($2\pi hc^2$)	$3.74177152466413 \times 10^{-8}$	$\text{W} \cdot \mu\text{m}^4/\text{m}^2$
$c_{1q\tilde{\nu}}$	First radiation constant ($2\pi c$)	$1.88365156730885 \times 10^{15}$	$\text{q} \cdot \mu\text{m}^3/(\text{s} \cdot \text{m}^2)$
$c_{2\tilde{\nu}}$	Second radiation constant (hc/k)	1.43877695998382	$\mu\text{m} \cdot \text{K}$
Radiation constants, ν expressed in spectral units of $[\text{Hz}]$			
$c_{1e\nu}$	First radiation constant ($2\pi h/c^2$)	$4.63227628074287 \times 10^{-50}$	$\text{J} \cdot \text{s}^3/\text{m}^2$
$c_{1q\nu}$	First radiation constant ($2\pi/c^2$)	$6.99098648422864 \times 10^{-17}$	s^2/m^2
$c_{2\nu}$	Second radiation constant (h/k)	$4.79924334848949 \times 10^{-11}$	$\text{s} \cdot \text{K}$
Wien's displacement law			
$w_{e\lambda}$	$10^6 hc/(a_5 k)$	2897.77212130396	$\mu\text{m} \cdot \text{K}$
$w_{q\lambda}$	$10^6 hc/(a_4 k)$	3669.70307542088	$\mu\text{m} \cdot \text{K}$
$w_{e\tilde{\nu}}$	$a_3 k/(100 hc)$	1.96099843866962	cm^{-1}/K
$w_{q\tilde{\nu}}$	$a_2 k/(100 hc)$	1.10762425613068	cm^{-1}/K
$w_{e\nu}$	$a_3 k/h$	$5.87892542062926 \times 10^{10}$	Hz/K
$w_{q\nu}$	$a_2 k/h$	$3.320573982858398 \times 10^{10}$	Hz/K
Stefan–Boltzmann law			
σ_e	Stefan–Boltzmann constant $\frac{2k^4\pi^5}{15c^2h^3}$	5.670373×10^{-8}	$\text{W}/(\text{m}^2 \cdot \text{K}^4)$
σ_q	Stefan–Boltzmann constant $\frac{4\zeta(3)k^3}{h^3c^2}$	1.5204×10^{15}	$\text{q}/(\text{s} \cdot \text{m}^2 \cdot \text{K}^3)$

Note: the q in the Units column is not an SI unit; it signifies quanta or photons.

Table A.4 Relative spectral efficiency for photopic and scotopic vision.^{4,5}

Wave-length λ	Photopic vision V_λ	Scotopic vision V'_λ	Wave-length λ	Photopic vision V_λ	Scotopic vision V'_λ
380	4.000×10^{-5}	5.890×10^{-4}	590	7.570×10^{-1}	6.550×10^{-2}
390	1.200×10^{-4}	2.209×10^{-3}	600	6.310×10^{-1}	3.325×10^{-2}
400	4.000×10^{-4}	9.290×10^{-3}	610	5.030×10^{-1}	1.593×10^{-2}
410	1.200×10^{-3}	3.484×10^{-2}	620	3.810×10^{-1}	7.370×10^{-3}
420	4.000×10^{-3}	9.660×10^{-2}	630	2.650×10^{-1}	3.335×10^{-3}
430	1.160×10^{-2}	1.998×10^{-1}	640	1.750×10^{-1}	1.497×10^{-3}
440	2.300×10^{-2}	3.281×10^{-1}	650	1.070×10^{-1}	6.770×10^{-4}
450	3.800×10^{-2}	4.550×10^{-1}	660	6.100×10^{-2}	3.129×10^{-4}
460	6.000×10^{-2}	5.672×10^{-1}	670	3.200×10^{-2}	1.480×10^{-4}
470	9.100×10^{-2}	6.756×10^{-1}	680	1.700×10^{-2}	7.160×10^{-5}
480	1.390×10^{-1}	7.930×10^{-1}	690	8.200×10^{-3}	3.533×10^{-5}
490	2.080×10^{-1}	9.040×10^{-1}	700	4.100×10^{-3}	1.780×10^{-5}
500	3.230×10^{-1}	9.817×10^{-1}	710	2.100×10^{-3}	9.140×10^{-6}
510	5.030×10^{-1}	9.966×10^{-1}	720	1.050×10^{-3}	4.780×10^{-6}
520	7.100×10^{-1}	9.352×10^{-1}	730	5.200×10^{-4}	2.546×10^{-6}
530	8.620×10^{-1}	8.110×10^{-1}	740	2.500×10^{-4}	1.379×10^{-6}
540	9.540×10^{-1}	6.497×10^{-1}	750	1.200×10^{-4}	7.600×10^{-7}
550	9.950×10^{-1}	4.808×10^{-1}	760	6.000×10^{-5}	4.250×10^{-7}
560	9.950×10^{-1}	3.288×10^{-1}	770	3.000×10^{-5}	2.413×10^{-7}
570	9.520×10^{-1}	2.076×10^{-1}	780	1.500×10^{-5}	1.390×10^{-7}
580	8.700×10^{-1}	1.212×10^{-1}			
nm			nm		

Table A.5 Infrared detector materials (1).^{7,8}

Parameter	Material			
	Ge	InSb	GaAs	InP
Bandgap E_g [eV]	0.742	0.235	1.519	1.4236
Refraction index	4	3.3	4	3.1
Varshni parameter A [eV/K]	0.48×10^{-3}	0.5405×10^{-3}	0.32×10^{-3}	0.363×10^{-3}
Varshni parameter B [K]	235	170	204	162
Electron mobility [$\text{cm}^2/(\text{V}\cdot\text{s})$] at 300 K	≤ 3900	$\leq 7.7 \times 10^4$	≤ 8500	≤ 5400
Hole mobility [$\text{cm}^2/(\text{V}\cdot\text{s})$] at 300 K	≤ 1900	≤ 850	≤ 400	≤ 200
Electron lifetime τ_e [s]	10^{-3}	10^{-10}	5×10^{-9}	10^{-8}
Hole lifetime τ_h [s]	10^{-3}	10^{-6}	2.5×10^{-7}	10^{-6}
Electron effective mass m_e^*/m_0	0.22	0.0135	0.067	0.0795
Hole effective mass m_h^*/m_0	0.33	0.43	0.45	
Lattice constant a_0 [\AA] at 300 K	5.6557	6.4794	5.65325	5.8697
Lattice constant temperature coefficient da_0/dT [$\text{\AA}/\text{K}$]		3.48×10^{-5}	3.88×10^{-5}	2.79×10^{-5}
Electron diffusion constant [cm^2/s]	≤ 100	$\leq 2\ 103$	≤ 200	≤ 130
Hole diffusion constant [cm^2/s]	≤ 50	≤ 22	≤ 10	≤ 5

Table A.6 Infrared detector materials (2).^{9 10}

Parameter	Material				
	Si	Ge	InSb	HgCdTe (x=0.22)	GaAs
Bandgap E_g at 300 K [eV]	1.107	0.67	0.163	0.102	1.35
Bandgap E_g temp. coeff. dE_0/dT [eV/K]	-2.3×10^{-4}	-3.7×10^{-4}	-2.8×10^{-4}	$+3.0 \times 10^{-4}$	-5×10^{-4}
Electron mobility μ_e [$\text{cm}^2/(\text{V}\cdot\text{s})$]	1900	3800	78000	3×10^5 at 77 K	8800
Hole mobility μ_h [$\text{cm}^2/(\text{V}\cdot\text{s})$]	500	1820	750	1×10^3 at 77 K	400
Electron mobility coeff. $d\mu_e/dT$ [$\text{cm}^2/(\text{V}\cdot\text{s}\cdot\text{K})$]	-2.6	-1.66	-1.6		-1
Hole mobility coeff. $d\mu_h/dT$ [$\text{cm}^2/(\text{V}\cdot\text{s}\cdot\text{K})$]	-2.3	-2.33	-2.1		-2.1
Electron diffusion constant [cm^2/s]	35	100			220
Hole diffusion constant [cm^2/s]	12.5	50			10
Intrinsic carrier concentration [cm^{-3}]	1.38×10^{10}	2.5×10^{13}		2×10^{13}	2×10^6
Electron effective mass m_e^*/m_0	1.1	0.55			
Hole effective mass m_h^*/m_0	0.56	0.37			
Refraction index near bandgap	3.42	4	4		3.65
Lattice constant a_0 [\AA] at 300 K	5.43072	5.65754	6.47877	6.47	5.65315
Absorption coefficient @ E_g α_{λ_c} [m^{-1}]			8×10^4		
Absorption coefficient α_0 [$\text{m}^{-1}\text{eV}^{-1/2}$]			1.9×10^6		
Bandgap E_g^Γ at 0 K [eV] from Piprek ¹⁰	4.34	0.8893	0.235		1.519
Varshni parameter A^Γ [eV/K] from Piprek ¹⁰	0.391×10^{-3}	0.6842×10^{-3}	0.32×10^{-3}		0.5405×10^{-3}
Varshni parameter B^Γ [K] from Piprek ¹⁰	125	398	170		204

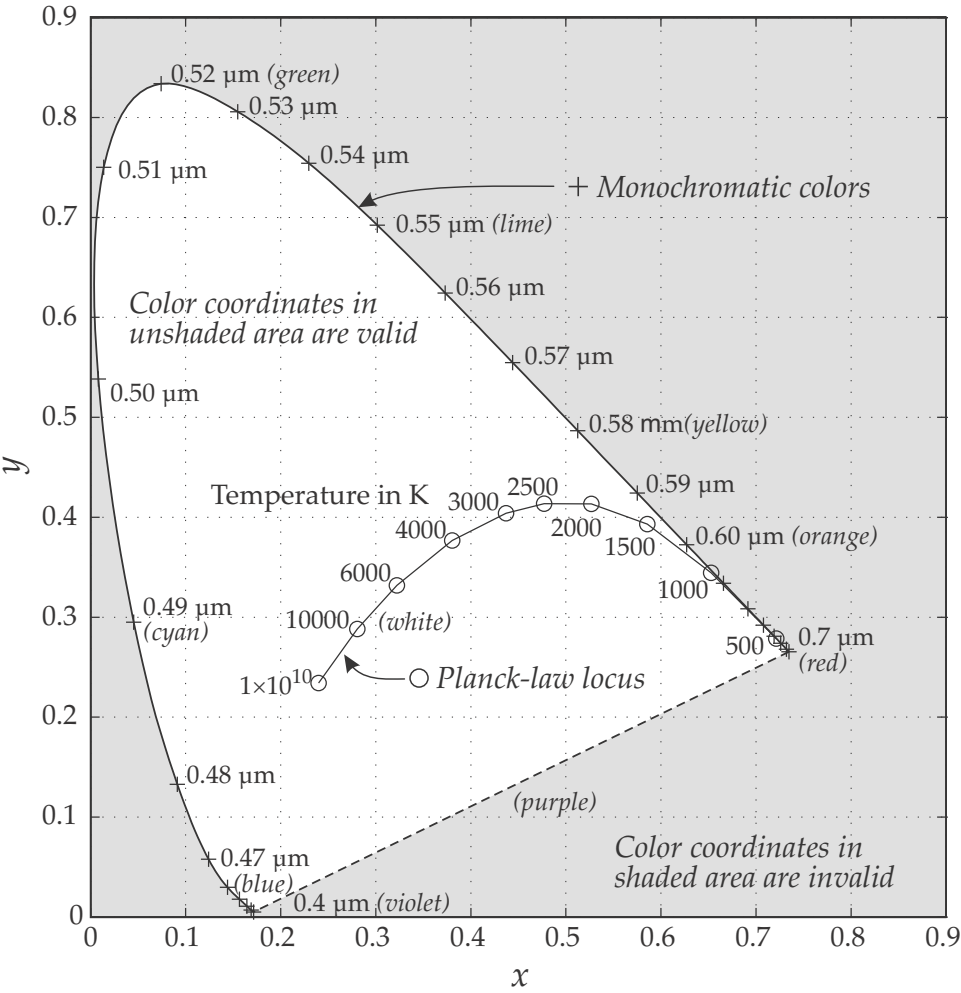


Figure A.1 CIE xy color chart. See Wikipedia⁶ for a color rendition.

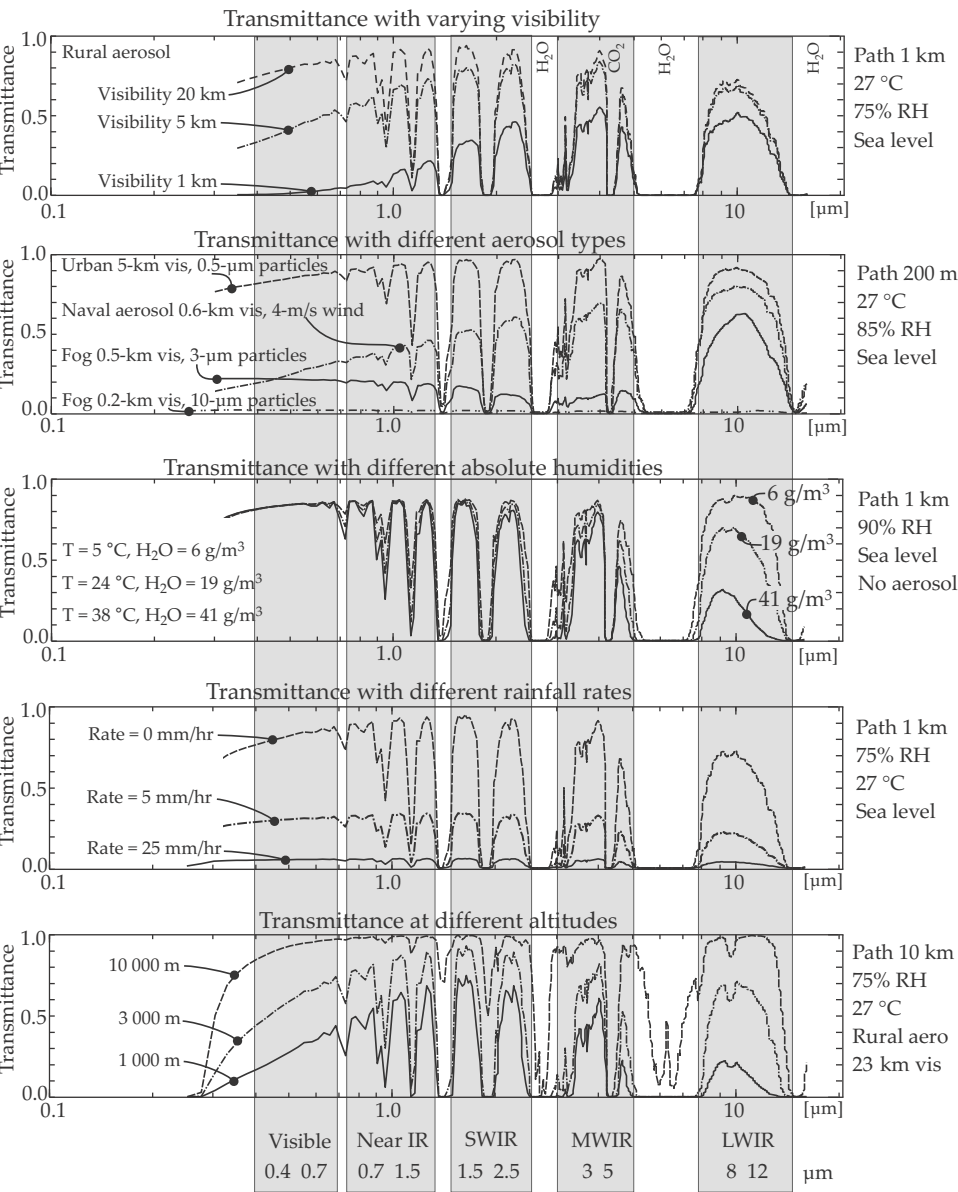


Figure A.2 Atmospheric transmittance for different climatic conditions.

Bibliography

- [1] Mohr, P. J., Taylor, B. N., and Newell, D. B., "CODATA recommended values of the fundamental physical constants: 2010," *Rev. Mod. Phys.* 84(4), 1527–1605 (2012) [doi: 10.1103/RevModPhys.84.1527].
- [2] SciPy, "SciPy Reference Guide: Constants (scipy.constants)," <http://docs.scipy.org/doc/scipy/reference/constants.html#codata2010>.
- [3] SpectralCalc, GATS Inc., "Radiance: Integrating the Planck Equation," http://www.spectralcalc.com/blackbody/integrate_planck.html.
- [4] Colour & Vision Research Laboratory, "Colour and Vision Database," <http://www.cvrl.org/index.htm>.
- [5] Pyradi team, "Pyradi data," <https://code.google.com/p/pyradi/source/browse>.
- [6] Wikipedia, "CIE 1931 color space," http://en.wikipedia.org/wiki/CIE_1931_color_space.
- [7] Levinshtein, M., Rumyantsev, S., and Shur, M., *Handbook Series for Semiconductors Parameters*, World Scientific (1996).
- [8] Vurgaftmann, I., Meyer, J. R., and Ram-Mohan, L. R., "Band Parameters for III-V Compound Semiconductors and their Alloys," *Journal of Applied Physics* 89(11), 5815–5875 (2001).
- [9] Dereniak, E. L. and Boreman, G. D., *Infrared Detectors and Systems*, John Wiley & Sons, New York (1996).
- [10] Piprek, J., *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*, Academic Press, San Diego, CA (2003).

INDEX

Index Terms

Links

Symbols

α , *see* absorptance

α , *see* absorption attenuation coefficient

β , *see* optical thickness

γ , *see* attenuation coefficient

Δf , *see* noise equivalent bandwidth

ϵ , *see* emissivity

η , *see* quantum efficiency

η_a , η_b , *see* image fill efficiency

η_s , *see* scanning efficiency

λ , *see* wavelength

λ_c , *see* cutoff wavelength

ν , *see* frequency, optical

$\tilde{\nu}$, *see* wavenumber

ρ , *see* reflectance

σ , *see* scattering attenuation coefficient

σ , *see* surface roughness

σ_e , *see* Stefan–Boltzmann constant

σ_q , *see* Stefan–Boltzmann constant

τ , *see* transmittance

Φ , *see* flux

ψ , *see* sun geometry factor

ω , *see* solid angle, geometric

Ω , *see* solid angle, projected

Ω_r , *see* field of regard

C_v , *see* contrast threshold

D , *see* pupil diameter

Index Terms

Links

D^* , *see* specific detectivity
 dA , *see* elemental area
 E , *see* irradiance
 E_g , *see* bandgap
 f , *see* electrical frequency
 f , *see* focal length
 F , *see* spatial view factor
 $F_\#$, *see* f -number
 f_r , *see* bidirectional reflection distribution function
 $f_{-3\text{ dB}}$, *see* bandwidth, -3 dB
 h , *see* Planck constant
 $h\nu$, *see* photon energy
 I , *see* intensity
 I_{ph} , *see* photocurrent
 I_{sat} , *see* reverse-bias-saturation current
 K_λ , *see* photopic efficacy
 k_f , *see* time-bandwidth product
 k_n , *see* noise equivalent bandwidth
 L , *see* radiance
 M , *see* exitance
 n , *see* index of refraction
 P_d , *see* probability of detection
 P_n , *see* probability of false detection
 q , *see* absolute humidity
 q , *see* quanta
 R , *see* responsivity
 R_V , *see* meteorological range
 S , *see* sensor response
 V_λ , *see* photopic vision
 V_λ , *see* scotopic vision
 Z_o , *see* detector preamplifier gain

Index Terms

Links

A

aberrations	232–235	
astigmatism	232	
chromatic	232	
comatic/coma	232	
distortion	235	
field curvature	235	
spherical	232	
absolute humidity	123	
absorptance		
attenuation coefficient	99	110
detector	242	
Kirchhoff's law	69	
material property	27	
absorption coefficient		
direct transition materials	177	
extrinsic semiconductor	178	183
free-carrier	177	
indirect transition materials	177	
intrinsic semiconductor	178	183
refractive index	176	
spectral	177	
typical curves	178	
Urbach tail	177	
advanced model, <i>see</i> lifecycle phases		
aerosols	112–116	
atmospheric transmittance	113	
land	112	
manmade	112	
maritime	112	
meteorological range	127	

Index Terms

Links

aerosols (<i>Cont.</i>)			
Mie scattering	116		
Rayleigh scattering	115		
scattering attenuation coefficient	128		
afocal optics	236	237	
aliasing	146	391	396–398
angle			
factor, <i>see</i> spatial view factor			
linear	27		
solid	28–35		
aperture stop	222	232	
approximation			
bandgap	139		
BRDF	81		
grey body	285		
layered atmosphere	104		
Planck law	64		
responsivity	415		
scattering	114		
scotopic efficiency spectral shape	46		
Seebeck coefficient	160		
solid angle	33	437	441
thin lens	221	225–227	
time bandwidth	150		
transmittance	108		
area			
clear aperture	242		
dimensional analysis	367		
elemental	19		
estimation of a flame	288–290	352	
example calculations	316	441–447	
pixel footprint	269		

<u>Index Terms</u>	<u>Links</u>	
area (<i>Cont.</i>)		
projected	28–35	
solid angle	28–35	366
spatial integral	407–409	
sun	316	
areance, <i>see</i> irradiance and exitance		
aspheric lens	237	
assumption management	11	
atmosphere	108–128	
absolute humidity	123	
aerosols	112	
attenuation	108	110
composition	108	
contrast transmittance	124–127	
definitions	109	
effect on image	268–272	
effective transmittance	107	
looking up/down	121	
meteorological range	127	
Mie scattering	116	
molecular		
absorption	111–112	
constituents	111	
transmittance	113	
overview	110	
path radiance	118–121	283
LWIR band	120	
MWIR band	119	
NIR band	118	
visual band	118	
radiative transfer codes	129	
Rayleigh scattering	115	

<u>Index Terms</u>	<u>Links</u>		
atmosphere (<i>Cont.</i>)			
relative humidity	123		
scattering	112	127	
scattering modes	114		
sky radiance	283	398–401	
standard profiles	109		
transmittance	113	382	
water vapor content	121		
windows	116		
LWIR band	117		
MWIR band	117		
NIR band	117		
visual band	116		
attenuation			
atmosphere	108		
coefficient	98–99		
avalanche detector	198		
B			
background	256		
background-limited operation	147	183	192
	205	211	
baffle	223		
band-limited noise	142		
bandgap	138–139		
Varshni approximation	139		
bandwidth			
– 3 dB	262		
Butterworth filter	263		
noise equivalent	262		
best practices	365–373		

<u>Index Terms</u>	<u>Links</u>		
bidirectional reflection distribution			
function (BRDF)	80–83		
Cook–Torrance model	82		
diffuse reflection	81		
measurements	83		
mirror reflection	81		
modeling approach	82		
Phong model	82		
reflection signatures	284		
specular reflective surface	327		
surface roughness	76		
blackbody			
aperture	75		
curves	68	69	72
definition	59		
emissivity	65		
Kirchhoff’s law	70		
laboratory instrument	59		
Lambertian source	41		
Planck’s law	60–62		
Stefan–Boltzmann law	63		
Wien’s displacement law	62		
Bloch functions	170		
bolometer	155–157		
construction	155		
noise	157		
responsivity	156		
Boltzmann probability distribution	58		
book website	xxv	411	
Bouguer’s law	98		
optical thickness	103		

Index Terms

Links

Bouguer's law (*Cont.*)

transmittance approximation 108

transmittance scaling 108

Bravais lattice 164

Bunsen burner flame case study

data analysis 350–355

instrument calibration 346–348

measurements 348–350

workflow 345–346

Butterworth filter 262

C

carrier lifetime 179

case study

Bunsen burner flame 344–355

cloud model 297–300

flame sensor 309–311

flame-area estimation 288

high-temperature flame measurement 295

infrared scene simulation 385–401

infrared sensor radiometry 337–344

laser rangefinder range equation 321–330

low-emissivity surface measurement 295

object appearance in an image 311–314

solar cell 315–321

sun-glint 302

temperature cross-over 300

thermal camera sensitivity 334–337

thermal imaging sensor model 330–334

thermally transparent paints 301

Cassegrain telescope 236 237

solid angle worked example 448

<u>Index Terms</u>	<u>Links</u>	
cavity	57	74
emissivity	74	
reflectance	74	
chief ray	224	230
cloud model case study		
measurement	297	
model	298–300	
relative signature contributions	300	
silver-lining factor	298	
worked example in Matlab®	451	
clutter	256	
CODATA constants, <i>see</i> constants		
cold finger	337	
cold shield	338	
design	342	
efficiency	341	342
collimator	238–239	
color		
coordinates	48–51	
worked example Python™	430	
normalization	48	
Planckian locus	49	
Ratio	291	398–401
sensitivity to source spectrum	49	
space, CIE 1931	48	
xy chart, CIE	381	
coma	234	
complex lens, <i>see</i> thick lens		
concept study, <i>see</i> lifecycle phases		
conduction band	168	
conductors	170	
energy bands	171	

<u>Index Terms</u>	<u>Links</u>	
configuration factor, <i>see</i> spatial view		
factor		
conservation of radiance	35–37	
constants		
CODATA	65	
mathematical	376	
physical	376	
Planck law	66	377
contrast		
difference	271	
inversion	300	
radiometric	272	
reduction	102	
signature	398–401	
threshold		
Koschmieder	127	
World Meteorological Organization	127	
transmittance	103	
atmosphere	124–127	
conversion		
radiometric to photometric	47	
spectral quantities	26	
convolution	265–267	
Cook–Torrance BRDF model	82	
\cos^3	32	449
\cos^4	33	
cryogenic coolers	185	
Joule–Thomson	185–186	
Stirling	186	
crystalline materials	163–179	
acceptor doping	172	
basis	164	

<u>Index Terms</u>	<u>Links</u>	
crystalline materials (<i>Cont.</i>)		
conductors	170	
donor doping	172	
energy bands	165–170	
insulators	170	
lattice	164	
n-type material	171	
p-type material	172	
pentavalent	171	
photon-electron interactions	174–176	
physical parameters	379–380	
semiconductors	170	
band structure	169–170	
intrinsic and extrinsic materials	171–174	
light absorption	176–178	
structure	164	
tetraivalent	170	171
cutoff wavelength	138	
D		
data analysis	292–295	
imaging-camera example	350–355	
workflow example	345–346	
definition study, <i>see</i> lifecycle phases		
design	2	3
prerequisites	3	
process	12	
review	6	
trade off	1	
detection		
probability	259	272
probability of false	260	

Index Terms

Links

detection (*Cont.*)

pulse	272–275	
pulse example calculation	436	
range	267–268	326
range example calculation	326–327	

detectivity	147–149	258	
specific	148	183	184
	205	258	

detector

avalanche	198	
conductivity	188	
configurations	140	
cooling	183–187	
gas/liquid cryogen	185	
radiative	185	
thermo-electric	185	
cutoff wavelength	138	
detection process	136–140	
detectivity, <i>see</i> detectivity		
dewar	185	
effective responsivity	148	
filter	338	
history	135–136	
intrinsic material	173	
material parameters	379–380	
noise	140–150	183
normalized spectral responsivity	140	243
peak responsivity	140	243
performance modeling	207–210	
photoconductive	179	187–193

<u>Index Terms</u>	<u>Links</u>		
detector (<i>Cont.</i>)			
photon	138–140		
detection process	179–183		
quantum efficiency	181–183		
photovoltaic	179	193–207	
preamplifier gain	243		
signal voltage	243		
spectral responsivity	182	243	
technology impact	210–212		
thermal	136–138	151–163	
wideband responsivity	261		
detector-limited operation	205	206	
development			
optronic sensor systems	385–386		
parallel activities	7		
phase, <i>see</i> lifecycle phases			
product	4		
development model, <i>see also</i> lifecycle			
phases			
dewar	185		
difference			
contrast	271–272		
noise equivalent temperature (NETD)	247	259	332–333
operator	19		
diffuse			
reflectance	76	81	
example visual spectra	50		
Phong BRDF model	82		
signature components	279–283		
reflectance, Phong BRDF	82		
shape factor, <i>see</i> spatial view factor			

<u>Index Terms</u>	<u>Links</u>		
dimensional analysis	367–368		
example	454	461	462
discrete ordinates	104		
distortion	234		
domain			
space	256		
time	256		
doped materials			
acceptor doping	172		
concentrations	174		
donor doping	172		
Duntley equations	101		
E			
effective, <i>see also</i> normalization			
detector responsivity	148		
mass			
electron	379		
hole	379		
transmittance	105–108		
example humid atmosphere	124		
example various sources	107		
scaling with range	108		
simulation application	393		
value normalization	261–262		
efficacy			
photopic	47		
scotopic	47		
total luminous	47		
efficiency			
cold shield	341		
image fill	331		

<u>Index Terms</u>	<u>Links</u>	
efficiency (<i>Cont.</i>)		
photopic	47	
quantum	139	
relative luminous	46	
human eye	47	378
scanning	330	
scotopic	47	
spectral shape approximation	46	
solar cell	318	
Einstein equation	180	
electrical frequency	141	
electro-optical system		
analysis		
example	309–364	
pyradi toolkit	411	
definition	14	
examples	15	
functions	221	
high-level design	15	
major components	14	
modeling and simulation	16	
multispectral	40	
simulation application	385–401	
electromagnetic		
radiation	20–22	
particle model	20	
wave model	20	
spectrum	21	
electron-hole pair	179	
elemental area	19	

Index Terms

Links

emissivity	65–74	
absorptivity	69	
atmosphere	120	121
blackbody	59	
cavity	74	
definitions	70	
directional	83–86	
example curves	85	
in nature	85	
gas radiator source	103	310
grey body	71	
Kirchhoff's law	69	
low	73	
measurement	295–296	
path radiance	101–103	
practical estimation	287–288	344–355
spectral	71	
hemispherical	84	
temporal variation	390	
thermally transparent paint	301	302
energy bands	165–170	
bandgap	165	
thermal carrier excitation	183	
conduction band	168	
Fermi level	166	168
Fermi–Dirac distribution	166	
interband transitions	174–176	
intraband transitions	174	175
orbitals	165	
photon-electron interactions	174–176	
semiconductor	169–170	
valence band	168	

<u>Index Terms</u>	<u>Links</u>	
energy bands (<i>Cont.</i>)		
wave model	166–169	
Bloch functions	170	
density of states	166	
wave function	167	
equivalent path length	99	
étendue, <i>see</i> throughput		
exitance	24	
Lambertian source	41–42	
luminous	23	
noise equivalent (NEM)	247	259
photon	23	60
Planck’s law	59–62	
temperature derivative	60–62	
radiant	23	38
relation to radiance	41	
source shape	44–45	
Stefan–Boltzmann law	63	
Wien’s displacement law	62	
experimental model, <i>see</i> lifecycle phases		
extended target	232	311–314
extinction coefficient, <i>see</i> attenuation		
coefficient		
extrinsic		
detector, <i>see</i> photon detector		
detector material	173	
eye spectral response	46	
F		
1/ <i>f</i> noise	142	145
photoconductive detectors	191	
power spectral density	145	

<u>Index Terms</u>	<u>Links</u>		
false alarm rate (FAR)	260		
pulse detection	272–275		
calculation in Matlab [®]	436		
calculation in Python [™]	436		
example calculation	273–275		
Fermi level	166	168	
Fermi–Dirac distribution	166	167	173
ferroelectric effect	157–158		
field			
angle	224	240	
curvature	234	235	
of regard	330		
of view (FOV)	227–232	240	
small angle	241		
stop	223	226–232	
figures of merit, <i>see</i> performance measures			
fill factor	317		
filter			
absorption	240		
antisolar	297		
Butterworth	262		
interference	240		
multi-spectral	39–41		
optical	240		
passband	240		
spectral	240		
function	413–415		
spectral response	223	240	
stopband	240		
transmittance	240		

<u>Index Terms</u>	<u>Links</u>	
flame		
area calculation in Matlab [®]	434	
Bunsen burner	344–355	
sensor	309–311	
worked example Matlab [®]	417	
worked example Python [™]	421	
temperature measurement	295	
fluctuation noise	146–147	
background flux	147	
signal flux	146	
flux	24	
collecting solid angle	230	231
Lambertian source	41	
luminous	23	
photon	23	
radiant	23	
system throughput	249	
transfer	35–41	70
geometrical construction	36	
lossless medium	37–38	
lossy medium	38	
multi-spectral	39–41	
radiative transfer equation	101	
worked example	448–451	
<i>f</i> -number (<i>f</i> /#)	229–230	
clear aperture area	242	
optics diameter	242	
focal		
length	224	
plane	223	224
folded optics	236	237
foreground	256	

<u>Index Terms</u>	<u>Links</u>	
frequency		
electrical	141	
optical	20	
relation to wavelength	20	
response		
photoconductive detector	190–191	
photovoltaic detector	202–203	
Fresnel reflectance	77–79	
gold surface	85	
full-width-half-maximum (FWHM) bandwidth	150	
G		
gaseous radiator	70–73	103–104
<i>see also</i> flame		
simulation	389	
generation–recombination (g-r) noise	144–145	
photoconductive detectors	191–192	
power spectral density	145	
rms noise current	144	
golden rules	365–373	
Gregorian telescope	236	237
grey body	71–73	
I		
I-V curve	196	
image	221	268
collimated, <i>see</i> collimator		
contrast	102–103	271
flux-collecting solid angle	230	
focal plane	223	

<u>Index Terms</u>	<u>Links</u>			
image (<i>Cont.</i>)				
modulation transfer function, <i>see</i>				
modulation transfer function				
(MTF)				
object appearance	311–314			
object relationship	225			
optical aberrations, <i>see</i> aberrations				
pixel irradiance	268–271			
pixels	268			
plane	223	227		246
field stop	227			
pupil	227			
vignetting	227			
point spread function, <i>see</i> point				
spread function (PSF)				
probability of detection	274			
ray tracing	225			
rendering	391–398			
resolved object	268–271			
simulation, <i>see</i> infrared scene simulation				
spatial sampling, <i>see</i> aliasing				
unresolved object	268–271			
image fill efficiency	331			
index of refraction	20			
atmosphere	97			
chromatic abberation	232			
complex	78	176		
Fresnel reflectance	78			
imaginary component	176			
metal	78			
numerical aperture (NA)	229			
real component	176			

<u>Index Terms</u>	<u>Links</u>	
index of refraction (<i>Cont.</i>)		
Snell's law	176	
wave equation	176	
industrialization, <i>see</i> lifecycle phases		
infinite conjugates	224	229
clear aperture	242	
collimator, <i>see</i> collimator		
<i>f</i> -number	229	
optics diameter	242	
infrared scene simulation	385–401	
application	387	
benefits	385	
image rendering, <i>see</i> rendering		
OSSIM	393	
radiometric accuracy	392	
rendering equation	393–396	
effective transmittance	394	
signature model	393	
spectral calculation	394	
spectral discretization	393	
wideband calculation	395	
scene model		
atmospheric attenuation	390	
geometry	388	
optical signature	388	
temperature	390	
temporal variation	390	
texture	390	
inhomogeneous medium	104	
insulators	170	

<u>Index Terms</u>	<u>Links</u>	
intensity	24	
Lambertian source	42	
luminous	23	46
photon	23	
radiant	23	38
interface electronics noise	146	
intrinsic carrier concentration	174	
intrinsic detector, <i>see</i> photon detector		
irradiance	24	
apparent	244	265
in an image	268–271	
<i>see also</i> object appearance in an image		
luminous	23	
noise equivalent (NEE)	246	258
<i>see also</i> laser rangefinder example		
photon	23	
pixel	268	
radiant	23	37
isolators		
energy bands	171	
J		
Johnson noise	142–143	
frequency spectrum	143	
interface electronics	146	
photoconductive detectors	191–193	
photovoltaic detectors	204	
power spectral density (PSD)	143	
K		
Kirchhoff's law	69	

<u>Index Terms</u>	<u>Links</u>	
knowledge management	386	
Koschmieder	127	
Kubelka–Munk theory	100	
L		
laboratory		
blackbody	59	75
collimator	238–239	
Lagrange invariant	249	
Lambertian source	41–42	
flux, exitance, radiance	41	
blackbody	41	
definition	41	
intensity	42	
projected solid angle	42	
reflectance	76	81
reflected sun radiance	87	
shape	44–45	
signature model	279–283	
spatial view factor	43	
view angle	42	
laser rangefinder		
detection range	326	
example calculation		
range equation	326–327	
signal-to-noise ratio (SNR)	274	
threshold-to-noise ratio (TNR)	274	
Lambertian reflective surface	323–325	
noise equivalent irradiance (NEE)	321	
range equation case study	321–330	
signal irradiance	322	
specular reflective surface	327–330	

<u>Index Terms</u>	<u>Links</u>			
laser rangefinder (<i>Cont.</i>)				
target optical cross section	324			
lifecycle phases	4–7			
light models	22			
light traps	76			
linear angle	27	28		
long-wave infrared (LWIR)	65			
atmospheric aerosol scattering	127–128			
atmospheric window	117			
contrast transmittance	125			
path radiance	120			
luminance	25	46–48		
photopic	47			
scotopic	47			
M				
marginal ray	224	229		230
material properties	27			
Matlab®	409			
measurement				
bidirectional reflection distribu-				
tion function (BRDF)	76	83		
cloud	297			
data analysis	292–295			
flame example	348–350			
instrument calibration	346–348			
linear angle	27			
spectroradiometer	287			
technical performance	255			
temperature	73	290–292		295
	354			

<u>Index Terms</u>	<u>Links</u>		
medium	14		
absorption attenuation coefficient	99		
atmosphere	108–128		
attenuation coefficient	99		
conducting	78		
discrete ordinates	104		
equivalent path length	99		
homogeneous	98		
index of refraction	20		
inhomogeneous	99	104–105	
lossless	37–38		
lossy	38		
optical	98–104		
optical thickness	103		
path radiance	99–103		
scattering attenuation coefficient	99		
transmittance	38	98	108
medium-wave infrared (MWIR)	65		
atmospheric aerosol scattering	127–128		
atmospheric window	117		
contrast transmittance	125		
path radiance	119		
mesopic vision	46		
meteorological range	127		
microbolometer	156–157		
Mie scattering	116		
minimum detectable temperature (MDT)	259		
minimum resolvable temperature (MRT)	259		
model	12		
atmospheric	129		
BRDF, <i>see</i> bidirectional reflection			
distribution function (BRDF)			

<u>Index Terms</u>	<u>Links</u>		
model (<i>Cont.</i>)			
cloud	297–300		
detector	207–210		
example	208		
discrete ordinates	104		
electromagnetic wave	20		
imaging sensor	240–245	337–344	
light	22		
photon particle	22		
photovoltaic detectors circuit	200		
signature	279–283		
solar cell	319–321		
solar irradiance	86		
source–medium–sensor	14		
validation	275		
modeling and simulation (M&S)	7	16	385–401
Modtran™			
description	129		
meteorological range	127		
visibility	127		
modulation transfer function (MTF)	236	260	
multi-spectral	39–41		
N			
near-infrared (NIR)	65		
atmospheric window	117		
path radiance	118		
Phong BRDF parameters	285		
noise	245–247	256	
bolometer	157		
considerations in imaging systems	146		

<u>Index Terms</u>	<u>Links</u>		
noise (<i>Cont.</i>)			
equivalent			
bandwidth	149–150	262	
exitance (NEM)	247	259	
irradiance (NEE)	246	258	
power (NEP)	147–149	246	258
radiance (NEL)	247	258	
reflectance (NER)	259		
target contrast (NETC)	335–337		
temperature difference (NETD)	247	259	332–333
$1/f$	145		
fluctuation	146–147		
generation–recombination (g-r)	144–145		
interface electronics	146		
Johnson	142–143		
Nyquist, <i>see</i> Johnson noise			
photoconductive detectors	191–193		
photovoltaic detectors	203–207		
physical processes	140		
power spectral density	141–142		
pyroelectric detector	159		
shot	143–144		
system	141		
temperature-fluctuation	145–146		
thermal, <i>see</i> Johnson noise			
thermoelectric detector	161		
time-bandwidth product	150		
normalization	261–263		
color coordinates	48		
effective value	261–262		
peak	262		

<u>Index Terms</u>	<u>Links</u>	
normalization (<i>Cont.</i>)		
spatial	29	261
weighted mapping	263	
normalized spectral responsivity	243	
n-type material	171	
electron concentration	174	
numerical aperture (NA)	229	230
O		
object		
appearance in an image	311–314	
worked example Python™	424	
resolved	268	
unresolved	268	
open-circuit operation	198	200
optics	223–236	
aberrations	232–235	
aperture	226	
aspheric lens	237	
axis	224	
chief ray	224	230
collimator	238	
conjugates	224	
elements	222–224	
field		
angle	224	
stop	226	230
flux collecting	230	
<i>f</i> -number	229	230
focal		
length	224	
plane	223	224

<u>Index Terms</u>	<u>Links</u>		
optics (<i>Cont.</i>)			
frequency	20		
infinite conjugates	224	229	
marginal ray	224	229	230
medium	97–104		
modulation transfer function (MTF)	236		
numerical aperture (NA)	229	230	
point spread function (PSF)	235		
power	223		
principal plane	224		
pupil	226–230		
ray tracing	225		
signature	279–292		
model	279–283		
rendering	387		
spectral filter	240		
stray light	227		
system	236		
afocal	236		
Cassegrain	236		
Gregorian	236		
refractive	236		
thick lens	225		
thickness	103		
thin-lens approximation	224	225	
transfer function (OTF)	236	260	
vignetting	227	238	
Optronics System Simulation (OSSIM)	393		
orbitals	165		

<u>Index Terms</u>	<u>Links</u>		
P			
paraxial approximation, <i>see</i> thin-lens approximation			
particle model	20		
passband	240		
path radiance	99–103	118–121	
Duntley equations	100		
emissivity	101–103		
Kubelka–Munk theory	100		
LWIR band	120		
MWIR band	119		
NIR band	118		
visual band	118		
Pauli’s exclusion principle	165	166	176
peak responsivity	243		
Peltier effect	151	186	
performance measures	10	255–261	
definition	256		
detectivity	258		
false alarm rate (FAR)	260		
minimum			
detectable temperature (MDT)	259		
resolvable temperature (MRT)	259		
modulation transfer function (MTF)	260		
noise equivalent			
exitance (NEM)	259		
irradiance (NEE)	258		
power (NEP)	258		
radiance (NEL)	258		
reflectance (NER)	259		
temperature difference (NETD)	259		

<u>Index Terms</u>	<u>Links</u>			
performance measures (<i>Cont.</i>)				
optical transfer function (OTF)	260			
point spread function (PSF)	260			
probability of detection	259			
probability of false detection	260			
role	255			
signal-to-clutter ratio (SCR)	257			
signal-to-noise ratio (SNR)	257			
specific detectivity	258			
Phong BRDF model	82			
phonon	175–177			
photoconductive detector	179	187		
bias circuitry	189–190			
conductivity	188			
frequency response	190–191			
geometry	189			
noise	191–193			
generation–recombination (gr)	192			
Johnson	192–193			
photoconductive gain	190			
quantum efficiency	187			
responsivity	189			
signal	187–189			
photocurrent	179	197	200	
	202	204	209	
photodiode, <i>see</i> photovoltaic detector				
photoemissive detector, <i>see</i> photon detector				
photometry	23	45–51		
units	45			
photon	22			
absorption	176–178			
absorption coefficient	177–178			

Index Terms

Links

photon (*Cont.*)

detector	138–140
noise, <i>see</i> noise	
operation	179
quantum efficiency	139
responsivity	139
electron interactions	174
energy	22
wave packet	22

photopic

efficacy	47
efficiency	47
luminance	47
relative spectral efficiency	378
vision	46

photovoltaic detector	179	193
background flux	204	
background-limited operation	205	
bias configurations	197–202	
circuit model	200	
open-circuit	200–202	
reverse	198–200	
short-circuit	202	
construction	194	
depletion region	194	
detector-limited operation		
open-circuit mode	206–207	
short-circuit mode	205–206	
diffusion current	197	204
energy diagrams	195	
frequency response	202–203	
I-V curve	196–197	

Index Terms

Links

photovoltaic detector (<i>Cont.</i>)	
noise	203–207
Johnson	204
shot	204
noise equivalent power (NEP)	204
optimal power transfer	202
photocurrent	204
p-n junction	194
quantum efficiency	196
resistance	204
responsivity	196
reverse-bias-saturation current	197
specific detectivity	205
thermally generated current	204
vs photoconductive detector	194
photovoltaic detectors	
energy bands	198
physical and mathematical constants	376
pixel	268
irradiance in an image	268–271
signal magnitude	268
Planck	
constant	22
exitance function Matlab [®]	412
exitance function Python [™]	412
law	57–65
constants	377
derivative exitance	60–62
exitance	60–62
integrated	63
maximum	62
summary	65

Index Terms

Links

Planck (<i>Cont.</i>)		
summation approximation	64	
radiator	59	65
Planckian locus	49	
plume	103	
effective transmittance	106	
surface radiator	104	
volume radiator	104	
p-n diode, <i>see</i> photovoltaic detector		
p-n junction	194	
point spread function (PSF)	235	260
point target	232	
Poisson statistics	144	
power spectral density (PSD)	141–142	
$1/f$ noise	142	145
band-limited noise	142	
combining spectra	149	
generation–recombination (g-r) noise	145	
Johnson noise	143	
shot noise	143	
temperature-fluctuation noise	145	
white noise	142	
principal plane	224	
probability of detection	259	
probability of false detection	260	
prototype, <i>see</i> lifecycle phases		
p-type material	172	
hole concentration	174	
pulse detection	272–275	
calculation in Matlab®	436	
calculation in Python™	436	
false alarm rate	272–275	

<u>Index Terms</u>	<u>Links</u>		
pupil	226–230		
diameter	229	230	
pyradi toolkit	411		
pyroelectric detector	157–159		
noise	159		
responsivity	159		
structure	158		
Python™	409		
Q			
quanta	20		
quantum efficiency	139	181	182
external	181		
anti-reflection coatings	182		
reflection	181		
internal	181		
photoconductive detector	187		
quantum well detector (QWIP), <i>see</i>			
photon detector			
R			
$1/R^2$ losses	311–314		
radiance	24		
atmospheric path	283		
basic	37		
conservation	35–37		
Lambertian source	41		
luminous (luminance)	25		
photon	25		
radiant	25		

<u>Index Terms</u>	<u>Links</u>	
radiance (<i>Cont.</i>)		
reflected		
ambient	283	
sky	283	
solar	283	
self-emitted	281	
signature model	279–283	
spatial invariance	36	
transfer	35–41	
transmitted background	283	
radiative transfer equation (RTE)	97	101
radiator		
gaseous	70	103
grey body	71	
Planck	59	65
selective	71	
surface	104	
thermal	285–292	
volume	104	
radiometer measurements		
atmospheric correction	267	
spectral radiance	287	
radiometric quantities	375	
radiometry	22	
definition	xxv	
nomenclature	23	
quantities	24	
techniques	255–276	
range equation	267	
solved in Python™	435	

<u>Index Terms</u>	<u>Links</u>			
ray				
chief	224			
marginal	224			
tracing	225			
Rayleigh scattering	115			
reductionism	xxiii			
reflectance				
bidirectional	76	80–83		
cavity	74			
diffuse	76	81		82
directional	75–83			
in nature	85			
Fresnel	77–79			
geometry	77			
high	73			
Lambertian	76	81		
material property	27			
mirror	81			
Snell’s law	77	79		405
specular	76	82		
refractive index, <i>see</i> index of refraction				
relative humidity (RH)	123			
relative luminous efficiency	46			
photopic	46			
scotopic	46			
rendering	387–398			
aliasing	391	396–398		
rasterization	391			
priority fill algorithm	391			
side-effects	393			
z-buffering	391			
super-sampling	392	396–398		

<u>Index Terms</u>	<u>Links</u>			
requirement allocation	8			
response				
eye	46			
filter	223	240		
frequency	150			
complex valued optical	236			
photoconductive detector	190–191			
photovoltaic detector	202–203			
impulse	235	260		
normalizing	140			
spatial frequency	260			
spectral weighting	106	244	263–264	
system	246			
thermal detector	136			
unlimited	161			
responsivity				
bolometer	156			
normalized	140	243		
peak	140	243		
photoconductive detector	189			
photon detector	139			
pyroelectric detector	159			
spectral	140	243		
thermal detector	136	152–154		
thermoelectric detector	161			
reverse-bias operation	198			
reverse-bias-saturation current	197	319		
review, <i>see</i> design				
root-mean-square (rms)	257			
S				
scanning efficiency	330			

<u>Index Terms</u>	<u>Links</u>	
scattering		
atmosphere	112	
aerosols	112	
attenuation coefficient	99	110
Mie	116	134
Rayleigh	115	134
scattering modes	114	
Schrödinger equation	169	
scotopic		
efficacy	47	
efficiency	47	
luminance	47	
relative spectral efficiency	378	
vision	46	
Seebeck coefficient	160	
selective radiator, <i>see</i> gaseous radiator		
semiconductors		
current flow	179	
carrier diffusion	179	
carrier drift	180	
charge mobility	180	
diffusion constant	180	
diffusion current	180	
diffusion current density	180	
drift current	180	
drift current density	180	
energy bands	171	
structure	169–170	
extrinsic materials	171	
concentrations	173	
examples	174	
Fermi energy level	173	

Index Terms

Links

semiconductors (<i>Cont.</i>)		
Fermi–Dirac distributions	173	
intrinsic materials	171	
concentrations	173	
examples	173	
Fermi energy level	173	
intrinsic carrier concentration	174	
light absorption	176–178	
material parameters	379–380	
Schrödinger equation	169	
silicon lattice	172	
wave equation	176	
sensor	14	
aperture stop	222	
field stop	223	
noise model	330–334	
optical		
elements	222	
model	240	
throughput	248–250	
optimization worked example Matlab®	459	
radiometric model	242–245	337–344
complex source	245	
detector signal	242	
source area variations	244	
signal calculations	242–245	
complex source	245	
detector	242	
source area variations	244	
solid angle		
field of view	230	
flux-collecting	230	

<u>Index Terms</u>	<u>Links</u>	
sensor (<i>Cont.</i>)		
spatial angles	230	
spectral		
filter	223	
response	223	243
stops/baffle	223	
terminology	221–223	
window	222	
worked example	450	
sharing	xxiii	
short-circuit operation	198	202
short-wave infrared (SWIR)	65	
shot noise	143–144	
interface electronics	146	
photovoltaic detectors	203	204
power spectral density	143	
signal	256	
reference planes	245	
electronics plane	246	
image plane	246	
object plane	246	
optics plane	246	
voltage	243	
signal-to-clutter ratio (SCR)	257	
signal-to-noise ratio (SNR)	257	
signature		
model	279–283	
atmospheric path radiance	283	
BRDF	284	
equation	281	
main contributors	280	
reflected ambient radiance	283	

Index Terms

Links

signature (<i>Cont.</i>)			
reflected sky radiance	283		
reflected solar radiance	283		
self-emitted radiance	281		
spatial properties	279		
terminology	282		
thermal radiator	285–292		
transmitted background radiance	283		
reflected vs emitted contribution	283		
rendering	387		
thermal radiation from common objects	65		
silicon detector	139		
simulation	385–401		
knowledge management	386		
validation	386		
sky radiance	283	398–401	
Snell’s law	79	176	403–405
solar cell analysis	315–321		
configuration	318		
experimental measurement	315		
model	319–321		
radiometry	317		
solid angles	316		
source areas	316		
solid angle	28–35		
approximation	33		
worked example Matlab®	441		
Cassegrain telescope example	448		
field of view	230		
flux collecting	230		
geometric	28		

Index Terms

Links

solid angle (*Cont.*)

cone 29

flat rectangular surface 32

projected 29 41

cone 31

flat rectangular surface 32

sphere 34

sensor 230

source area 366

source 14

gaseous 103

Lambertian 41–42

shape, *see* Lambertian source, shape

space domain 256

spatial

integral 38 407

calculation in Matlab[®] 437

view factor 43

specific detectivity 148 183 184

205 258

photon-noise-limited thermal detector 161

specification hierarchy 8

specifications 8–10

spectral

band (NIR, MWIR, SWIR, LWIR) 65

calculations 264–267

convolution 265–267

detector function 415–417

in Matlab[®] 415

in Python[™] 416

domains 25

Index Terms

Links

spectral (*Cont.*)

emissivity	71		
measurement	288		
filter	223	240	
filter function	413–415		
in Matlab®	413		
in Python™	414		
filtering	39		
integral	407		
integration, summation	26		
mismatch	264		
quantities	25		
conversion	26		
density	25		
response			
eye	46		
filter	240		
photon and thermal detectors	138		
sensor	223	246	
responsivity	243		
weighting	106	244	263–264
spectroradiometer	287		
specular reflectance	76	82	
Stefan–Boltzmann law	63		
stopband	240		
subsystem	2		
sun	86		
area	316		
geometry factor	87		
glint	302		
reflected radiance	283	398–401	
surface radiator	104		

<u>Index Terms</u>	<u>Links</u>
surface roughness	75
scale	76
system	2
acceptance, <i>see</i> lifecycle phases	
context	1
engineering	2
noise	245–247
performance measures	255–261
segment, <i>see</i> subsystem	
source–medium–sensor model	242–245
V-chart	8
T	
target	
extended	268
point	268
technical performance measure (TPM)	10
telescope	
Cassegrain	236
Gregorian	236
temperature	
apparent	73
cross-over	300
estimation of a flame	290–292
minimum detectable	259
minimum resolvable	259
noncontact measurement	73
radiation	73
temperature-fluctuation noise	145–146
flux	146
power spectral density	145

<u>Index Terms</u>	<u>Links</u>	
thermal detector	136–138	151–163
bolometer	155–157	
conceptual model	152	
noise, <i>see</i> noise		
overview	151–152	
photon-noise-limited	161–163	
noise equivalent temperature		
difference (NETD)	163	
specific detectivity	161–163	
pyroelectric	157–159	
responsivity	136	152–154
temperature-fluctuation-noise-limited	163	
noise equivalent temperature		
difference (NETD)	163	
specific detectivity	163	
thermoelectric	159–161	
thermal imager		
sensitivity	334–337	
sensor model	330–334	
assumptions	330	
electronic parameters	330	
example calculation	333	
flux on the detector	337–339	
focused optics	339–342	
noise	331–333	
out-of-focus optics	342–344	
thermal radiator, <i>see</i> grey body, <i>see</i>		
Planck radiator		
white point	49	92
thermal radiator model	285–292	
area estimation	288	
emissivity estimation	287	

<u>Index Terms</u>	<u>Links</u>	
thermal radiator model (<i>Cont.</i>)		
process	287	
temperature estimation	290	
thermally transparent paint	301	
thermocouple		
equation	161	
gas measurement	354	
thermodynamic equilibrium	57	
thermoelectric coolers	186–187	
thermoelectric detector	159–161	
layout	160	
noise	161	
responsivity	161	
thick lens	225–227	
thin-lens approximation	221	224–227
throughput	248–250	
time domain	256	
time-bandwidth product	150	
transfer function		
modulation (MTF)	236	260
optical (OTF)	236	260
transmittance		
atmospheric windows	116	
LWIR band	117	
MWIR band	117	
NIR band	117	
visual band	116	
background	283	
Bouguer's law	98	
contrast	103	
effective	105	
filter	240	

<u>Index Terms</u>	<u>Links</u>	
transmittance (<i>Cont.</i>)		
homogeneous medium	98	
inhomogeneous medium	99	
material property	27	
medium	38	98
range	108	
two-flux Kubelka–Munk	100	
U		
up/down atmospheric radiance	121	
Urbach tail	177	
V		
valence band	168	
validation	275	386
value system	11	
Varshni approximation	139	
V-chart, <i>see</i> system		
vignetting	222	227
collimator beam	238	239
control of	228	
in practical design	341–342	
visibility, <i>see</i> meteorological range		
vision		
mesopic	46	
photopic	46	
scotopic	46	
visual spectral band		
atmospheric window	116	
contrast transmittance	125	
path radiance	118	

Index Terms

Links

volume radiator 104

W

wave model

 electronic 166

 Bloch functions 170

 field strength 176

 velocity 176

 wave equation 176

 light 20

wave packet 22

wavefront 20

wavelength 20

 cutoff 138

 relation to frequency 20

 relation to wavenumber 26

 spectral density conversion 26

wavenumber 25

website xxv 411

white noise 142

white point 49 92

Wien's displacement law 62–63 67

Z

zero field angle, *see* optical axis